




Acoustic Events Detection and Classification for Urban DAS Data with Supervised Learning

Yichen Zhong¹, Chen Gu^{*1} , Germán A. Prieto² , Michael Fehler³ , Peng Wu¹, Zhi Yuan¹, Weiwei Xu⁴, Shanhui Xu⁴, and Xinzheng Lu¹ 

Abstract

A key aspect of improving disaster prevention and mitigation in sustainable smart cities is to increase the sensory capabilities of existing communication infrastructure, providing reliable information for urban management in emergency situations. Distributed acoustic sensing (DAS) is an advanced technology suitable for this application because it has a wide range of applications, including urban environmental awareness, structural health monitoring, and disaster warning. In this study, the field test data are measured by a DAS array deployed along the edge of the Guye area in the city of Tangshan in China, where the 1976 Tangshan earthquake occurred. We analyzed the vibrations from natural and artificial acoustic sources across both the space and frequency domains and revealed various characteristics of the sources. Subsequently, a deep learning-based method was developed for multiple acoustic source detection and classification, including earthquake, vibrator vehicle, traffic flow, and industrial production. The training dataset was created using this acquisition of DAS field data, which was annotated using the label transfer method proposed in this article. Then, typical acoustic events are classified and extracted from DAS data in the space–frequency domain. The proposed source identification scheme enables real-time monitoring of routine urban activities with long-distance coverage and high accuracy, as well as detection of abnormal events. In addition, we can use this method to expand the range of recognized classes or apply it to special datasets. Our study shows a great value in improving the ability of urban environmental perception and hazard information analysis. It also holds potential for earthquake detection, site-effects studies, and anomaly detection in urban environments.

Cite this article as Zhong, Y., C. Gu, G. A. Prieto, M. Fehler, P. Wu, Z. Yuan, W. Xu, S. Xu, and X. Lu (2025). Acoustic Events Detection and Classification for Urban DAS Data with Supervised Learning, *Seismol. Res. Lett.* **96**, 3112–3131, doi: [10.1785/0220240222](https://doi.org/10.1785/0220240222).

Introduction

Smart cities are a new trend in modern urban development, aiming to utilize advanced technology and innovative solutions to improve urban sustainability, efficiency, and quality of human life (Giffinger *et al.*, 2007). In this process, intelligent sensing plays a crucial role (Hancke *et al.*, 2013). Existing research uses geophones (Li *et al.*, 2021) and video streaming (Fedorov *et al.*, 2019) for traffic flow monitoring and fire detection (Talaat and ZainEldin, 2023). Rashidi *et al.* (2011) utilize body motion sensors and hot/cold water usage sensors to track human activities and evaluate their health performance. These sensors are often deployed as a massive number of nodes, requiring a large-scale installation on-site to achieve full urban monitoring coverage. However, with the continuous expansion of the urban scale, node-based sensing systems may face limited coverage and high maintenance cost. The distributed acoustic sensing (DAS) technology provides a new approach

to address these problems. By extensively utilizing pre-existing telecommunication fiber optics in the city, it is possible to monitor human activities and infrastructure health (Liu *et al.*, 2023). Compared to traditional node-based sensors, distributed fiber-optic systems have the advantages of wide coverage area, low maintenance costs, and more timely response, which enable the management to have a more timely and comprehensive understanding of the city's operational status.

1. Department of Civil Engineering, Tsinghua University, Beijing, China,  <https://orcid.org/0000-0003-0003-8992> (CG);  <https://orcid.org/0000-0002-3313-7420> (XL); 2. Departamento de Geociencias, Universidad Nacional de Colombia, Bogotá, Colombia,  <https://orcid.org/0000-0001-8538-7379> (GAP); 3. Earth Resources Laboratory, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A.,  <https://orcid.org/0000-0002-8814-5495> (MF); 4. Institute of Geophysics, China Earthquake Administration, Beijing, China

*Corresponding author: guchch@tsinghua.edu.cn

© Seismological Society of America

DAS technology utilizing phase-optical time-domain reflectometry (Φ -OTDR) facilitates distributed measurement of the strain rate over distances spanning several tens of kilometers with a higher spatial resolution (Shatalin *et al.*, 1998). A DAS system typically consists of an interrogator unit and optical fiber cables. The interrogator emits continuous short-pulse laser signals (light) into the cables. The laser signal undergoes reflection and scattering through the fiber material and then returns to the interrogator. When the fiber experiences vibration or acoustic disturbances, it alters the light scattering process in the optic fiber. By analyzing the scattered light returned to the interrogator using Rayleigh backscattered theory physical quantities such as strain rate induced by vibrations can be measured. DAS can achieve sampling rates ranging from less than 0.001 Hz–50 kHz.

The advantageous features of DAS technology facilitate wide application in geological science, including geological exploration, earthquake source studies, ocean wave observations and mining (Williams *et al.*, 2019; Li *et al.*, 2022). In addition, it also has numerous applications in urban settings, such as structural health monitoring, critical infrastructure security, and surveillance of urban anthropogenic activity (Zhan, 2019; Lindsey and Martin, 2021; Chen *et al.*, 2023; Cheng *et al.*, 2023). The work of Kowarik *et al.* (2020) and Wang *et al.* (2022) has shown that DAS systems can achieve conventional traffic flow monitoring with good accuracy. Lindsey *et al.* (2020) and Wang *et al.* (2021) analyzed the impact of the COVID-19 pandemic on human activities using DAS data. Cunningham *et al.* (2023) found DAS system records mine activities and local and regional blasting events very well.

A key aspect of using such a data-rich technology is to be able to automatically extract information because doing this manually is not feasible. Various machine learning methods have been used on the DAS data to extract the signature of the acoustic source and its identification. Wang *et al.* (2020) used XGBoost to recognize active source events. Chiang *et al.* (2023) proposed a 1D CNN architecture for DAS signal classification and it performed well with good accuracy. Yuan *et al.* (2024) achieved high spatial resolution vehicle identification using a spatial deconvolution U-Net model. A seismic signal feature extraction model called PhaseNet-DAS was proposed based on the phase association method of the U-Net and Gaussian mixture model, which has the potential to use the DAS technique to enhance earthquake monitoring (Zhu *et al.*, 2023).

These studies demonstrate that machine learning methods are very efficient at processing massive DAS data and are effective in extracting acoustic source features. Most studies have focused on extracting acoustic features from temporal or spectral domains using single-channel data, whereas DAS data can capture the spatial distribution of acoustic sources, revealing the source propagation patterns across space and exhibiting pronounced signal features. Classification in the space–frequency domain is robust and computationally efficient.

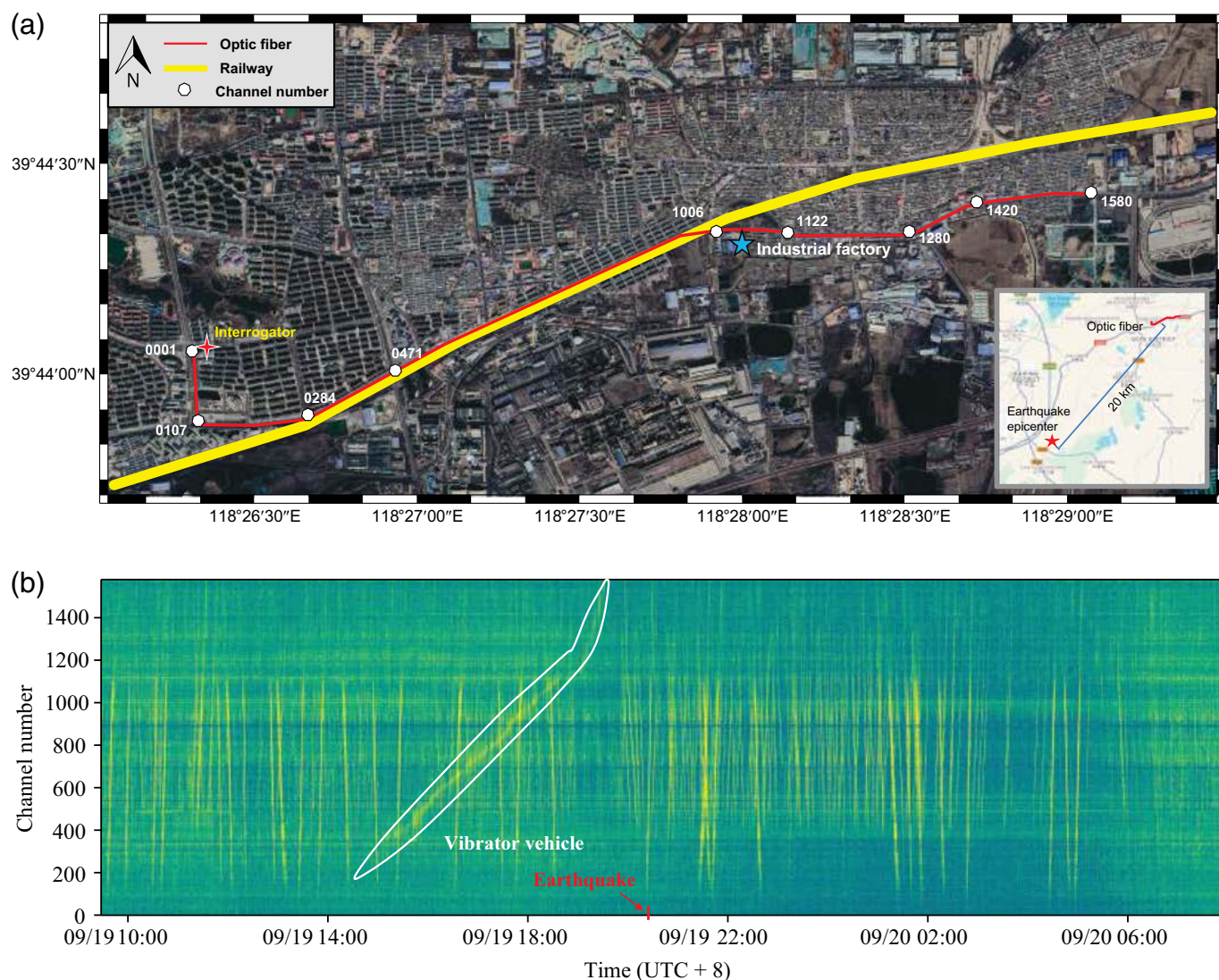
In this study, we transformed DAS data into space–frequency spectrograms using the short-time Fourier transform (STFT) and trained a YOLOv8 model to achieve the real-time detection and classification of acoustic events. To obtain the training dataset, we proposed a label transfer method allowing manual data annotation in the spatiotemporal domain, and the annotated data were transformed into the space–frequency domain for model training. We also conducted urban-scale DAS field observation experiments, collecting rich DAS data that encompassed human activities and natural events such as traffic flows, factory production, and seismic disasters. Subsequently, we analyzed the signal features of different event types. These DAS data served as the foundation for training our model. Compared to the similar detection methods, our approach effectively utilizes the space–frequency pattern of signals, greatly facilitating detection efficiency. This study indicates that DAS technology exhibits promising prospects for both routine urban monitoring and anomaly detection. It enables the observation of tiny vibrations and the extraction of time–space–frequency features in the urban environment, which can be used for structural monitoring and disaster prevention.

Data and Methods

Field data

We measured field test data using an optical fiber array deployed in the Guye area in the city of Tangshan in China, as shown in Figure 1a, which is 43 km away from the epicenter of the M_w 7.5 Tangshan earthquake that occurred in 1976. The total length of the optical fiber is about 6.45 km and it traverses the central urban area of Guye from west to east. The optical fiber passes through residential zones, near industrial factories and major urban roadways, and it is also near a railway line. During the experiment, an earthquake occurred near this area. Consequently, the acquired data contain a wide variety of signals, including human activities and natural events. In this experiment, the Silixa iDAS device was used as the interrogator unit. The interrogator was placed at the western end of the optical fiber, as shown in Figure 1a. We used a 10 m gauge length and channel spacing of 4.08 m for DAS data acquisition. The total measured range is ~6446 m, with 1580 channels. We used Global Positioning System to determine the actual spatial distribution of the fiber-optic cable. The spatial position of the optical fiber, as well as the locations of several channels, are shown in Figure 1a.

Figure 1b shows an example DAS record (about 22 hr) that includes multiple signals corresponding to both human and natural sources. We analyzed the collected DAS data and identified many interesting acoustic signal events. Hammering tests were conducted along the optical fiber route with a vibrator vehicle and provided a constant manual active source. The occurrence of an earthquake is indicated with a red arrow on 19 September at around 20:22 (UTC + 8). The DAS signal



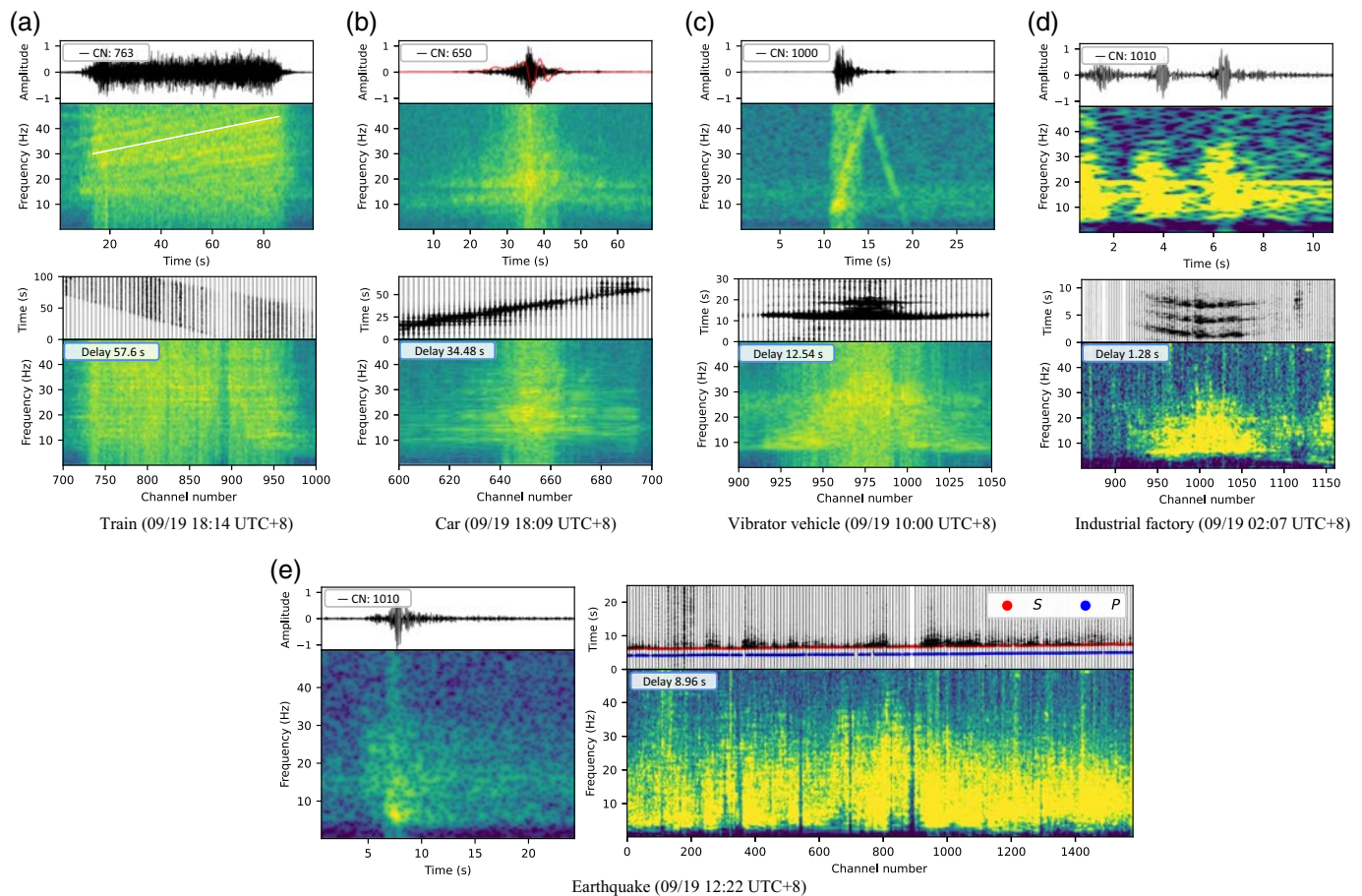
due to the earthquake is not obvious in Figure 1b because the ground motions caused by the earthquake are much weaker compared to those generated by the trains near the fibers. However, when we zoom-in on the earthquake occurrence time, the seismic waves become clearly visible, as shown in Figure 2e. In addition, the signal due to the movement of a vibrator vehicle is outlined with white lines, starting from the afternoon of 19 September and continuing into the evening. To understand the features of the various acoustic events, time–frequency spectrograms and space–frequency spectrograms of typical events were created separately. The results of the analysis are presented in Figure 2 and are discussed subsequently.

Train source

This is due to the action of moving harmonic forces on the steel rails (Sheng *et al.*, 2004). Figure 2a shows the data analysis of a train source. The train signal exhibits clear periodicity due to the similar excitation characteristics of multiple carriages. The train source generates a wideband and wide-range acoustic

Figure 1. (a) Layout of the fiber-optic cable (red line) in the Guye Area, China. The bold yellow line represents the railway line. The white dots mark the locations of some distributed acoustic sensing (DAS) channels. The blue star indicates the industrial factory with distinguishable signal features. The inset in the map shows the relative location of the optic fiber (red line) and a DAS detected earthquake (red star). (b) A section of the DAS data records containing several typical acoustic signals. The red marker indicates the occurrence of an earthquake, where the location is marked with a red star in the inset map in panel (a), and the white lines mark the signal of a vibrator vehicle. The earthquake signal is not prominent in this figure because the ground motions caused by the earthquake are much weaker than those generated by the trains near the fibers. Detailed earthquake waveforms can be seen in a zoom-in window near the earthquake occurrence in Figure 2e. The color version of this figure is available only in the electronic edition.

signal. When the train moves at a constant speed, the excited frequency bands remain stable, resulting in horizontal lines in the time–frequency spectrogram of a single channel. However,



when the train speed changes, a frequency-shift effect occurs, causing the previously horizontal lines to become sloped (e.g., the white line shown in Fig. 2a). Furthermore, the higher the frequency, the greater the slope variation. Zhang *et al.* (2022) use these features of multichannel signals to estimate the train speed.

Car source

The car signal is characterized by a series of short-duration (~20–30 s), high-energy signals. Its waveforms and spectrograms are displayed in Figure 2b. Although vehicles also generate wideband excitations, their primary frequency components concentrate around 5–15 Hz, and different vehicle types exhibit different frequency characteristics. After applying a low-pass filter, the vehicle-induced vibration signal exhibits a noticeable undulation, as seen by the red line on the waveform graph. This corresponds to quasi-static deformations of the road surface caused by the vehicle (Chiang *et al.*, 2023; Corera *et al.*, 2023).

Vibrator vehicle source

During the observation period, a seismic vibrator vehicle was deployed to carry out active-source signal tests, which also increased the diversity of signal sources. The ground vibrations caused by the vibrator vehicle are significantly stronger, and

Figure 2. DAS records of typical events recorded during our DAS development. (a–d) Sources of train, car, vibrator vehicle, and industrial factory, with each figure displaying four panels: a single channel number (CN) waveform, a time–frequency spectrogram of the waveform, a multichannel waveform and a space–frequency spectrogram from top to bottom. The delay time label indicates the timestamp of the space–frequency spectrogram relative to the starting time (0 s). The red line in the first panel of car represents the quasi-static waveform after low-pass filtering. (e) The earthquake recorded by the DAS. The earthquake panels on the left represent the single-channel waveform and spectrogram, and the panels on the right represent the multichannel waveform and space–frequency spectrogram. The arrival times of *P* waves and *S* waves of the earthquake are marked with red and blue dots, respectively, in the figures. The short-time Fourier transform (STFT) time windows are set to 2.56 s, with an overlap of 0.9 between adjacent time windows. The color version of this figure is available only in the electronic edition.

their propagation range is much larger than that of conventional vehicles. From the spectrum analysis in Figure 2c, the vibrations generated by the seismic source vehicle exhibit a unique feature. The frequency shows a linearly increasing trend over time, and the signal intensity distribution across various frequency bands received by DAS channels near the source location appears relatively uniform.

Industrial factory source

There is a factory that manufactures refractory materials near our observation region and we annotated its location with a blue star in Figure 1a. The factory operates during the day, and the acoustic signals generated by the mechanical equipment are recorded by the DAS system. To better illustrate the wavefield characteristics of the factory source, the waveforms and spectrograms are plotted in Figure 2d. This source has a frequency ranging from ~5 to 30 Hz, with an influence radius about 600 m. From the space–frequency spectrogram, high-frequency signals are concentrated near the source location, gradually attenuate toward both ends away from the source. Thus, the signal features in the space–frequency domain show a triangular shape.

Earthquake source

An earthquake was recorded during the DAS observation, as detailed in Figure 2e. We employed PhaseNet-DAS (Zhu et al., 2023) to determine the arrival times of *P* waves and *S* waves for each channel, which are marked on the graph. The energy of the *S* wave is stronger than that of the *P* wave. We also observed significant amplitude variations across different locations of the DAS array, which may be attributed to the difference in the near-surface properties of the region (Aki, 1993). Assuming a uniform *P*-wave velocity of 5 km/s (Ma et al., 2023), we used a 2D grid-search method, ignoring depth information, to locate the epicenter ~20 km southwest of the DAS array. This location is indicated by a red star in the inset of Figure 1a.

This analysis of seismic source signals illustrates that different acoustic events exhibit distinct features in the space–frequency domain. It is feasible to identify events using wavefield information in the time–space–frequency domain. Thus, we propose a machine learning-based acoustic event detection and classification method for urban DAS data, which has good efficiency and enables long-range and real-time monitoring of urban operational status and anomaly detection.

Data processing methods

In the data processing section, the multichannel DAS data were transformed into space–frequency domain spectrograms using the STFT. Subsequently, a YOLO model was trained to identify acoustic events within these space–frequency spectrograms. The dataset employed for training the model was prepared through the label transfer method (described in the following [Datasets and Label Transfer](#) section).

We first downsampled the raw DAS data to 100 Hz, which can cover the frequency range of typical acoustic events while maintaining computational efficiency. Then, to obtain the space–frequency spectrogram, we applied the STFT to compute the time–frequency representation for the DAS data (Oppenheim, 1999):

$$X(\tau, \omega) = \sum_{t=-\infty}^{\infty} x(t) w(t - \tau) e^{-j\omega t}, \quad (1)$$

in which $X(\tau, \omega)$ denotes the spectrogram of a time-series waveform $x(t)$, τ , and ω are, respectively, time and angular frequency, $w(t)$ denotes the time window, and j is the imaginary unit. After processing the DAS data by STFT, the original 2D waveform (spatiotemporal waveform) was transformed into 3D data, incorporating time, space, and frequency, as indicated by the translucent blue cuboid in Figure 3a. Based on this 3D data structure, we proposed a workflow as shown in Figure 3b. The detailed event classification process is as follows:

- Step a, we first selected DAS data of any length and converted it into 3D data blocks using STFT. Note that the data length must exceed the STFT window length. As shown in the red dashed box on the left side of Figure 3b.
- Step b, we then extracted data from different channels at the same time to form a space–frequency spectrogram. The space–frequency spectrogram was then used as an input data for the neural network model. In this study, the YOLO-v8 detection model (Jocher et al., 2023) was used to detect different acoustic events in the space–frequency domain. YOLO-v8 is a version of the YOLO model series known for its speed, accuracy, and ease of use. This is the first study to apply this model to object detection and classification using DAS data. The YOLO-v8 framework simplifies the training process through high-level encapsulation (Ultralytics, 2024). We do not need to manually design complex model architectures and it offers a more efficient training experience. The loss function of YOLOv8 is a combination of complete intersection over union loss and distribution focal loss (DFL) for bounding box regression, and binary cross-entropy (BCE) loss for classification.
- Step c, the time slice spectrograms were sequentially fed into the detector in chronological order. It annotated the positions of events for each timeframe using bounding boxes and provide labels for different event types and that is step d.

Through the proposed method, the positions and types of acoustic events at different time were inferred. Thus, the real-time multiobject detection using DAS data were achieved.

Datasets and Label Transfer

The label transfer method was proposed for preparing the training dataset for the YOLO model. In this section, we selected an 8-min data segment (19 September 2022 18:48–18:56) as a representative case. First, multichannel waveforms were visualized to generate a DAS image in the space–time domain. Manual annotation was performed on the DAS image to extract spatiotemporal event labels, with 36% of the annotations completed by researchers and the remaining 64% by subcontracted human workers. By incorporating the frequency characteristics of distinct events, space–time–frequency annotations were derived for each event and stored as a structured label list. Subsequently, each channel of the raw DAS data

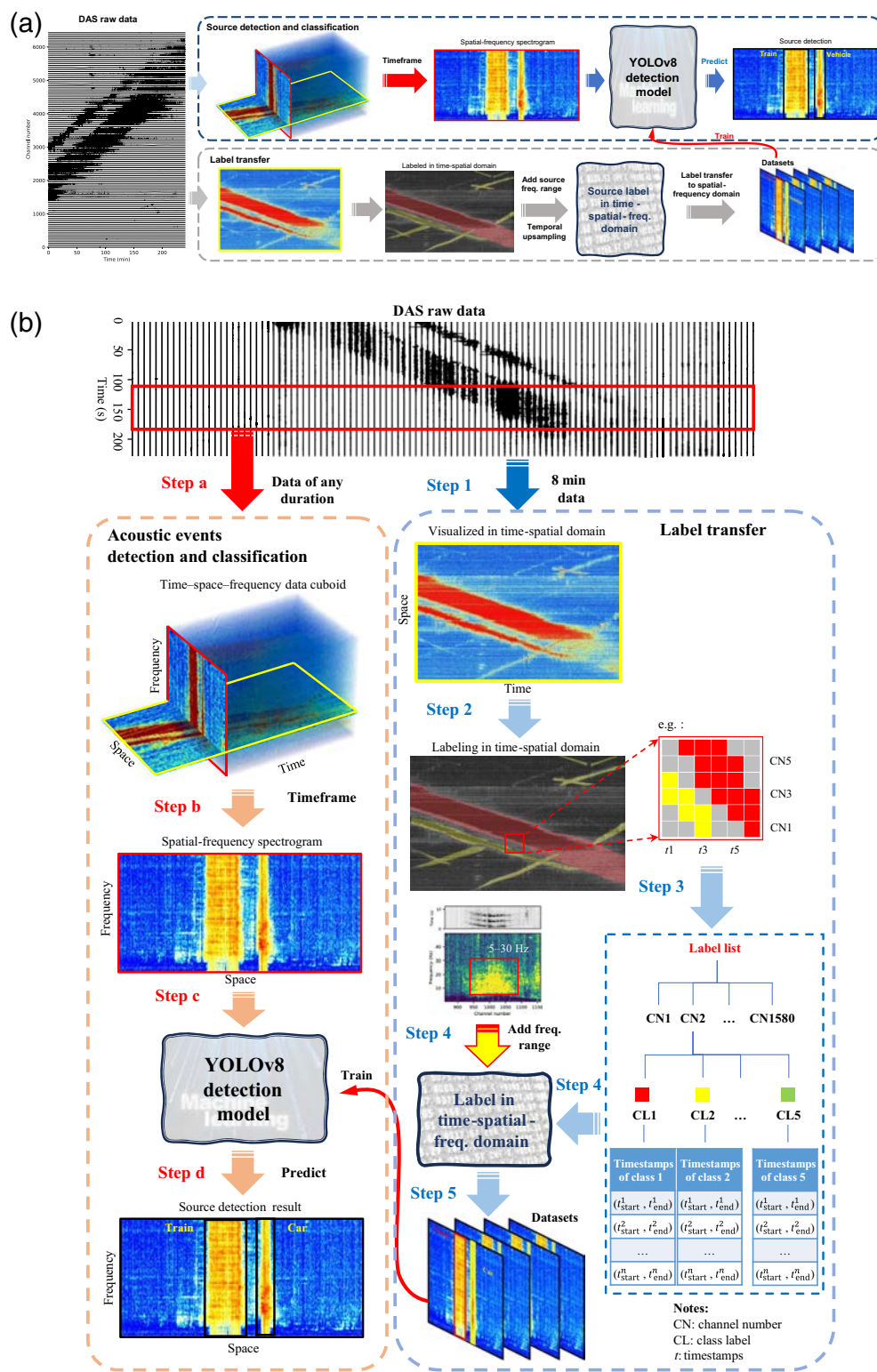


Figure 3. (a) The time–space–frequency data cuboid, and slicing it in time–space domain and space–frequency domain, respectively. (b) The proposed methodology for acoustic event detection and classification. The red left dashed rectangle above represents the data processing flow for acoustic event detection and classification, and the blue right dashed rectangle represents the processing flow of the label transfer method. The color version of this figure is available only in the electronic edition.

underwent STFT, expanding the dataset into the space–time–frequency domain. Data slices corresponding to identical timestamps were extracted to generate space–frequency spectrograms. Finally, annotations from the label list were mapped to the current timestamp to assemble the dataset.

The label transfer method allows us to annotate a single DAS spatiotemporal image and then generate many annotated samples in the space–frequency domain. For example, an 8-min segment of data at 100 Hz, after applying the STFT with a time window w_{len} of 2.56 s (256 sample points) and a timestep T_{step} of 2.56 s (256 sample points), produced 187 spectrograms. This method is briefly described as the series of blue arrows shown in Figure 3b. The detailed process is described subsequently:

- Step 1, as shown in the blue dashed box on the right side of Figure 3b. We selected an 8-min length section of data to create the training dataset using the label transfer method. The data length was adjusted according to the actual task requirements. For this study, the reasons for selecting an 8-min duration are detailed in the [Label transfer data processing length](#) section. Then, the downsampled DAS raw data were visualized in the spatio-temporal domain. The 8-min section of raw data we resampled at 10 Hz, giving 4800 points in the time dimension. Taking into account the 1580 channels in the spatial dimension, we obtained a 2D matrix of size 1580×4800 in the

spatiotemporal domain that is visualized as an image. The vertical and horizontal axis of this image, respectively, represent channel number and time.

- Step 2, acoustic events were annotated in this spatiotemporal image. As shown in Figure 3b, similar to the annotation method used in semantic segmentation, we label each pixel in the image to distinguish different categories. We utilized relevant annotation tools (Wada, 2018) to draw polygons to annotate different parts of the image, rapidly generating the corresponding pixel-level labels. During the annotation process, when encountering overlapping sources, our usual approach is that the signal source with stronger energy overrides the signal source with weaker energy. In addition, we have manually included annotated spectrograms to improve the differentiation of overlapping sources. This allows for multiple labels for signals at the same time and space, helping the model distinguish between different sources at the same time and location. Please refer to the [Handling overlapping signal sources](#) section for details.
- Step 3, according to the pixel level labels, the start time t_{start} and the end time t_{end} of different acoustic events were obtained for any given channel, as shown in Figure 3b. Then these timestamps of the event were recorded in a label list at the channel level.
- Step 4, we obtained the label information for all channels at any given time through the label list. However, frequency information was also required for the identification and classification task. As the example in Figure 2, the frequency characteristics of different acoustic events are distinct. The acoustic sources of train, car, and vibrator vehicle have very broad-frequency bands, which cover the entire valid frequency range (around 0–50 Hz). The factory and the earthquake have relatively narrow-frequency bands, which are 5–35 Hz and below 40 Hz, respectively. In addition, we conducted a detailed power spectral density (PSD) analysis specifically for car events, which demonstrated that the frequency range for each type of event remains relatively stable. For further details, please refer to the [Power spectral density analysis of car events](#) section.
- The step 5 was to transfer the labels from the spatiotemporal domain to the space–frequency domain. As mentioned earlier, the 8 min section data generated 187 spectrogram images. The central time T_i of each image can be calculated by timestep and time window with equation (2), which is required for further processing,

$$T_i = T_s + w_{\text{len}}/2 + T_{\text{step}} \times i, \quad (2)$$

in which T_s represents the start time of the 8-min segment of DAS data, i means the index of spectrogram image. For the spectrogram at time T_i , the horizontal range (space axis) and the center position of the bounding box were determined according to the label list. In addition, based on the analysis

TABLE 1

Number of Labels for Each Class in the Guye Dataset

Class	Train	Car	Vibrator Vehicle	Industrial Factory	Earthquake
Number	428	1688	146	148	76

of the frequency characteristics of different acoustic events mentioned earlier, the vertical range (frequency axis) and center position of the bounding box were determined. Once the position and range of the bounding box are determined in the spectrogram images, the label transfer work is completed.

Combined with the object detection and classification method proposed in this article, it can help researchers effortlessly achieve custom object detection and classification tasks with a large volume of DAS data. In this study, we used a portion of the entire DAS raw data and created a spectrogram dataset for model training through the label transfer method; and the labeled dataset is split with 85% used for training and 15% for validation. The number of samples for each class label is shown in Table 1.

Results and Discussion

The YOLOv8 model was trained on the Guye DAS dataset created by the label transfer method. The batch size was set to 16, and we trained on this dataset for 600 epochs with initial learning rate 0.01, incorporating an early stopping mechanism (stopping training early because no improvement was observed in the last 50 epochs). We trained and tested five different configurations of YOLOv8 model, including nano (YOLOv8-n), small (YOLOv8-s), medium (YOLOv8-m), large (YOLOv8-l), and extra-large (YOLOv8-x) size, each with increasing number of parameters, as shown in Table 2. The proposed model was trained, validated, and evaluated on a workstation, which has an Intel i9-13900K central processing unit (CPU) and Nvidia RTX4090 graphic processing units (GPU).

The configuration and evaluation for each model are shown in Table 2, with $\text{Speed}_{\text{CPU}}$ and $\text{Speed}_{\text{GPU}}$ being tested on the workstation mentioned earlier. In this table, P represents precision, R represents recall, and mAP_{50} refers to the mean average precision when the intersection over union threshold is set at 50%. It evaluates how well the predicted bounding boxes overlap with the ground-truth boxes, considering a detection correct if the overlap is at least 50%, and mAP_{50-95} means an overlap ranging from 50% to 90% of this metric. Figure 4 illustrates the overall trend of the performance metrics of the proposed model, including loss, precision, and recall, and the comprehensive explanation of those metrics can be found in the official documents (Ultralytics, 2025). The training loss is slightly greater than the validation loss during training.

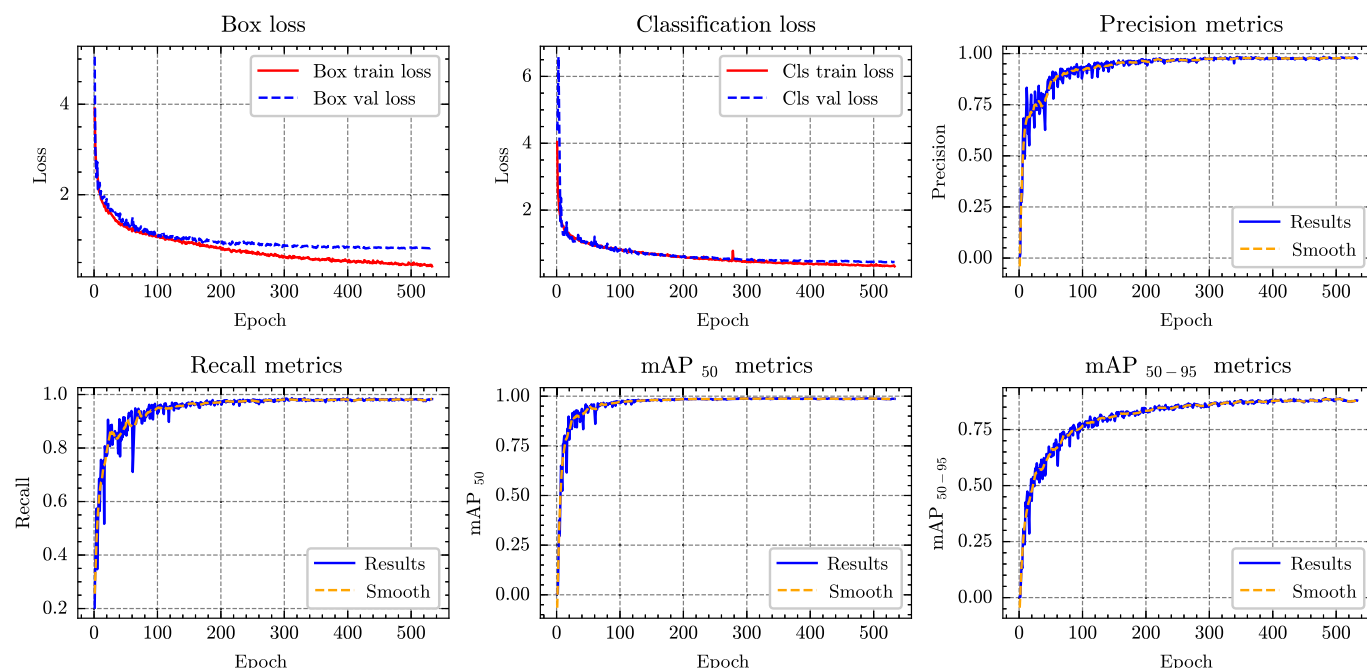


Figure 4. Training results of the acoustic event detection and classification model. The color version of this figure is available

only in the electronic edition.

TABLE 2

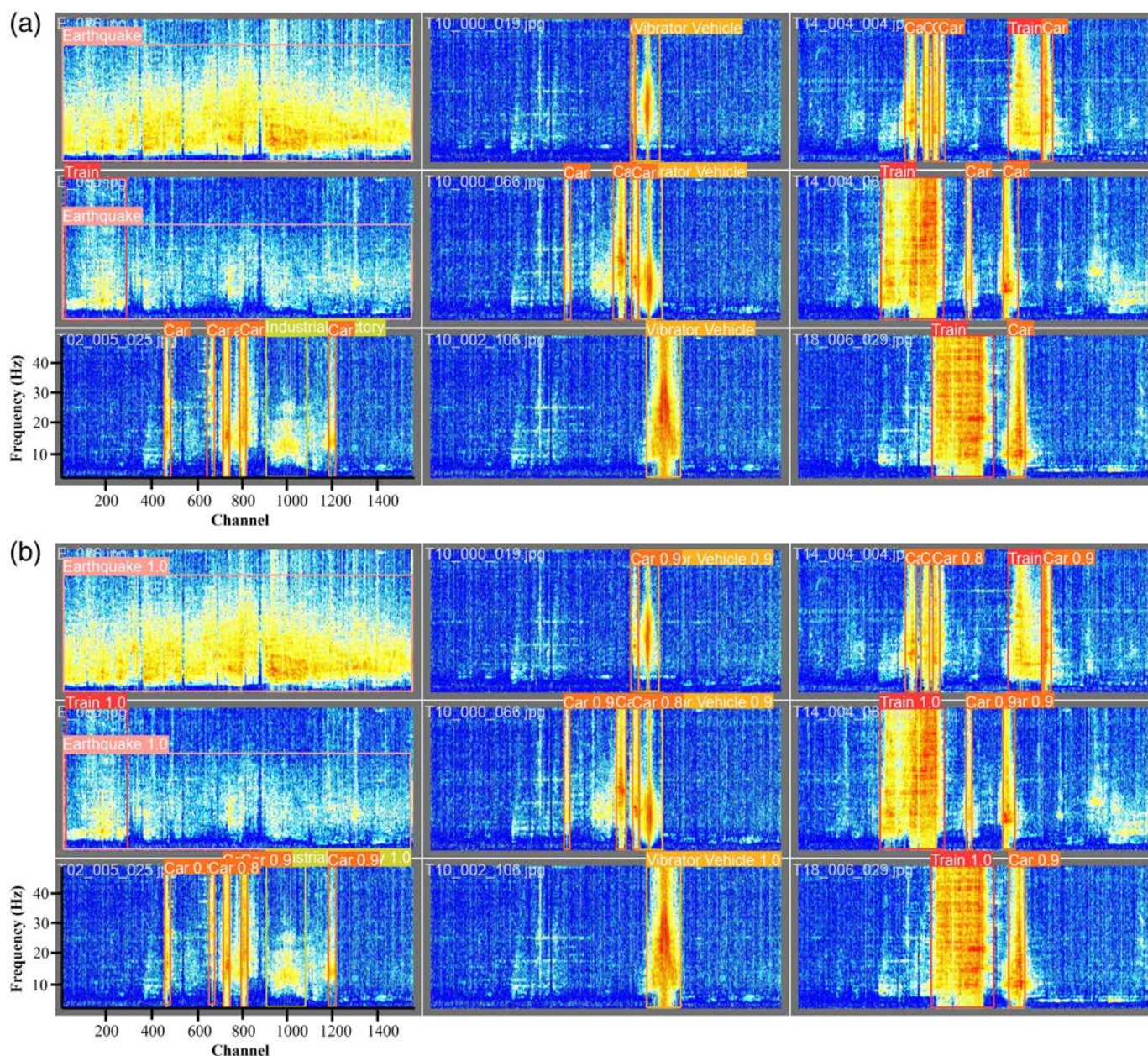
Performance of the YOLOv8 Model on the Distributed Acoustic Sensing (DAS) Dataset

Model	<i>P</i>	<i>R</i>	mAP ₅₀	mAP _{50–95}	Speed _{CPU}	Speed _{GPU}	Params	GFLOPs
YOLOv8-n	0.984	0.978	0.986	0.876	6.7 ms	4.1 ms	3006623	8.1
YOLOv8-s	0.990	0.983	0.988	0.892	19.4 ms	4.8 ms	11127519	28.4
YOLOv8-m	0.986	0.983	0.987	0.891	54.9 ms	5.5 ms	25842655	75.7
YOLOv8-l	0.986	0.987	0.987	0.900	86.9 ms	6.1 ms	43610463	164.8
YOLOv8-x	0.983	0.985	0.988	0.896	133.6 ms	6.9 ms	68128383	257.4

However, both the training loss and validation loss decrease steadily throughout the training process, with no signs of overfitting. Figure 5 shows the test results of this detection and classification model.

The earlier results demonstrate that the proposed method performs well in acoustic source detection and classification on the DAS dataset. All class detection accuracy of multiple configuration models is above 98%. For the more stringent evaluation metric mAP_{50–95}, it can also reach 0.900. The confusion matrix in Figure 6 illustrates the model recognition results for the six classes. The model performs excellent recognition results for trains, earthquakes, industry factory, and vibrator vehicles, which approaches 100%. The detection rate for the car class is also very high. However, there are 24 background labels identified as cars, which may be due to some vehicles being missed during manual labeling of the DAS data, resulting in the model

recognizing these targets without corresponding labels. The proposed method exhibits high accuracy in detecting both common urban human activities, such as traffic sources and uncommon acoustic sources such as earthquakes or vibrator vehicles. We also tested this method on the entire DAS dataset; detailed information can be found in the [Performance on the whole dataset](#) section. In addition, it is observed that when the model scale is up to YOLOv8-s or larger, the increase in the number of parameters has a relatively marginal improvement on the model's accuracy. All class mAP_{50–95} of YOLOv8-x are about 0.896, whereas YOLOv8-l achieve 0.900. It may be attributed to the fact that when acoustic waveform data collected by the DAS system are transformed into the space–frequency domain, the wavefield features of typical events become more distinct and easier to differentiate. Therefore, the detection and classification problem are relatively simple, requiring no need for large-parameter models



for feature extraction. This further confirms that the proposed method has considerable potential to identify a wide range of acoustic events.

We also performed efficiency tests on both CPU and GPU platforms. On the GPU platform, the processing time for a single space–frequency spectrogram by each model is within the millisecond range. Therefore, there is no need to consider the model size, and it is preferable to select the high-accuracy YOLOv8-l for DAS data processing. On the CPU platform, the computation speed decreases significantly with increasing model size, ranging from 6.7 ms for YOLOv8-n to 133.6 ms for YOLOv8-x. Therefore, we need to trade-off between model accuracy and processing speed. We recommended running YOLOv8-s on the CPU platform for the classification and detection of DAS data because it offered excellent accuracy and an acceptable processing speed.

Figure 5. The performance of the proposed method for acoustic event detection and classification in the test dataset. (a) The label of the DAS datasets created using Guye field data, and (b) the prediction results of the method. The red boxes indicate trains, orange boxes indicate cars, yellow boxes indicate vibrator vehicles, pink boxes indicate earthquakes, and green boxes indicate industrial factory sources. The digits within the filled rectangles in panel (b) represent the confidence scores for each predicted class. Each image represent a single space–frequency spectrogram for a 2.56 s time window. The color version of this figure is available only in the electronic edition.

In addition, the performance of the supervised learning model requires large and diverse data for training and validation. We used a commercial AI trainer service to achieve original labels in time–space domain for the six classes we defined, before applying our label transfer method. Because time and

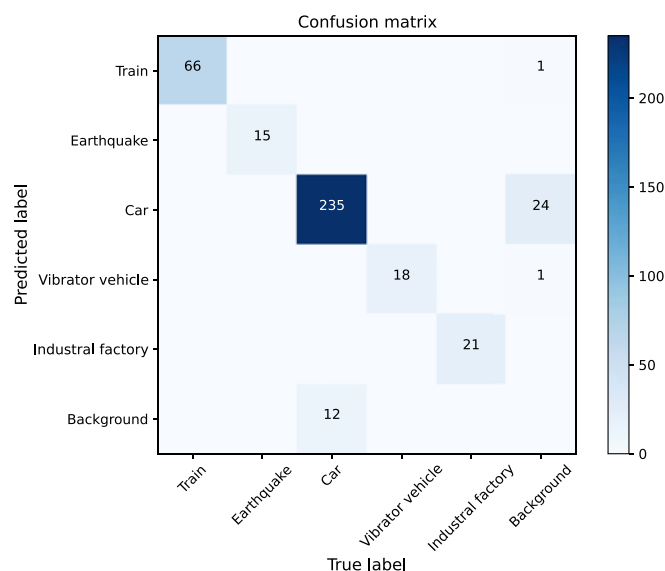


Figure 6. Confusion matrix of predicted results. The color version of this figure is available only in the electronic edition.

space scales expand in the future, more classes of acoustic events may be discovered. To address this, we will employ a commercial service to annotate new classes and expand the categories of our datasets. Our model, which has shown strong generalization performance across other datasets (see the [Model generalization analysis](#) section) can serve as a pretrained model for transfer learning, paving the way for future improvements and adaptations to an evergrowing range of acoustic events.

Conclusions

We have proposed a long-range and high-accuracy object detection and classification method, which takes advantage of DAS acquisitions and advanced deep learning model YOLOv8. This method can be widely applied to routine human activity monitoring, including monitoring of traffic and industrial production, and also detection of hazardous events in smart cities, such as earthquakes. It provides a new approach for modern smart urban sensing. The proposed method utilizes the characteristics of acoustic events in the space–frequency domain, greatly improving the accuracy and efficiency of identifying acoustic signals. The identification accuracy of all classes can be up to 99%, and the more stringent evaluation metric mAP_{50-95} also can be up to 0.900. The method also exhibits excellent operational efficiency. When running on a GPU platform, the processing speed per frame is in the millisecond range, which is totally capable of meeting the requirements of routine real-time monitoring. In addition, the proposed label transfer method can be easily utilized by relevant researchers to enrich detection categories or applied to specific datasets. The proposed smart city monitoring method has potential applications in urban environmental perception and hazard information

analysis and can be used to improve the safety and sustainability of urban environments.

Data and Resources

The code will be available at https://github.com/ZHONG-YICHEN/DAS_ML.git. The distributed acoustic sensing (DAS) data used in this study can be obtained from the corresponding author on a reasonable request. To evaluate the generalization ability of the model proposed in this study, we conducted tests on multiple datasets, including the foresee dataset published on PubDAS (Spica *et al.*, 2023) and the Southern California Earthquake Data Center (SCEDC) earthquake data AWS Public dataset (Southern California Earthquake Data Center [SCEDC], 2013) (<https://scedc.caltech.edu/data/getstarted-pds.html>) (Zhu *et al.*, 2023). All websites were last accessed in May 2024.

Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

Acknowledgments

The research is supported by the National Key Research and Development Program of China (Grant Number 2022YFC3003602), National Natural Science Foundation of China (Grant Number 52238011), and National Natural Science Foundation of China (Grant Number 42274062). The authors acknowledge the Institute of Geophysics at the China Earthquake Administration for their data support.

References

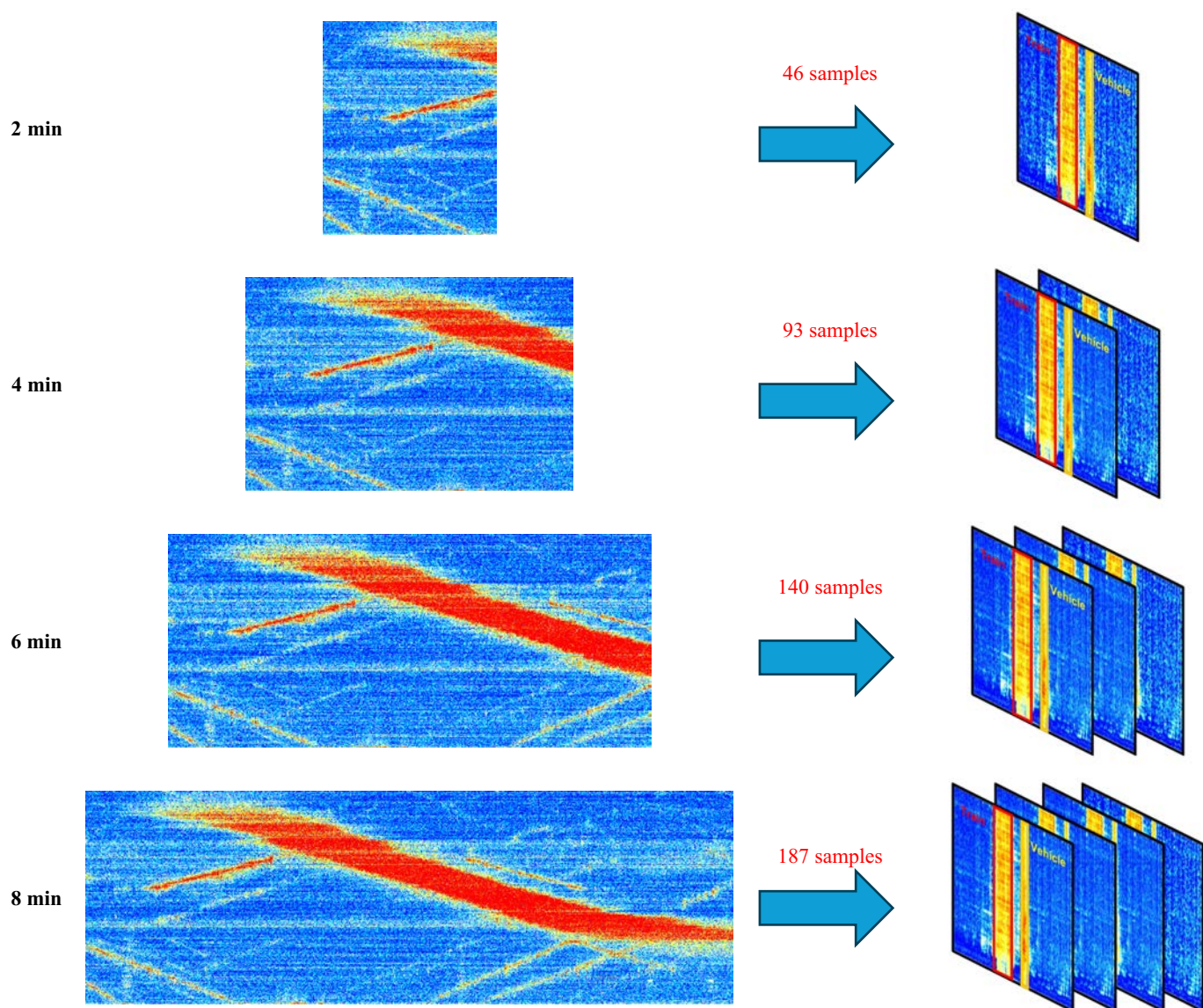
- Aki, K. (1993). Local site effects on weak and strong ground motion, *Tectonophysics* **218**, nos. 1/3, 93–111.
- Chen, S., J. Wang, C. Zhang, M. Li, N. Li, H. Wu, Y. Liu, W. Peng, and Y. Song (2023). Marine structural health monitoring with optical fiber sensors: A review, *Sensors* **23**, no. 4, 1877, doi: [10.3390/s23041877](https://doi.org/10.3390/s23041877).
- Cheng, F., J. B. Ajo-Franklin, A. Nayak, V. R. Tribaldos, R. Mellors, P. Dobson, and the Imperial Valley Dark Fiber Team (2023). Using dark fiber and distributed acoustic sensing to characterize a geothermal system in the imperial valley, southern california, *J. Geophys. Res.* **128**, no. 3, e2022JB025240, doi: [10.1029/2022JB025240](https://doi.org/10.1029/2022JB025240).
- Chiang, C.-Y., M. Jaber, K. K. Chai, and J. Loo (2023). Distributed acoustic sensor systems for vehicle detection and classification, *IEEE Access* **11**, 31,293–31,303, doi: [10.1109/ACCESS.2023.3260780](https://doi.org/10.1109/ACCESS.2023.3260780).
- Corera, I., E. Piñeiro, J. Navallas, M. Sagues, and A. Loayssa (2023). Long-range traffic monitoring based on pulse-compression distributed acoustic sensing and advanced vehicle tracking and classification algorithm, *Sensors* **23**, no. 6, 3127, doi: [10.3390/s23063127](https://doi.org/10.3390/s23063127).
- Cunningham, E., N. Lord, D. Fratta, A. Chavarria, C. Thurber, and H. Wang (2023). Three-dimensional distributed acoustic sensing at the Sanford Underground Research Facility, *Geophysics* **88**, no. 6, WC209–WC220, doi: [10.1364/AO.422385](https://doi.org/10.1364/AO.422385).
- Fedorov, A., K. Nikolskaia, S. Ivanov, V. Shepelev, and A. Minbaleev (2019). Traffic flow estimation with data from a video surveillance camera, *J. Big Data* **6**, no. 1, 73, doi: [10.1186/s40537-019-0234-z](https://doi.org/10.1186/s40537-019-0234-z).
- Giffinger, R., C. Fertner, H. Kramar, R. Kalasek, N. Milanovic, and E. Meijers (2007). *Smart Cities - Ranking of European Medium-Sized*

- Cities, Vol. 2007, Centre of Regional Science at the Vienna University of Technology (UT), Vienna, Austria, 1–12.
- Hancke, G. P., B. D. C. e. Silva, and G. P. Hancke Jr. (2013). The role of advanced sensing in smart cities, *Sensors* **13**, no. 1, 393–425, doi: [10.3390/s130100393](https://doi.org/10.3390/s130100393).
- Jocher, G., A. Chaurasia, and J. Qiu (2023). Ultralytics YOLO, available at <https://github.com/ultralytics/ultralytics> (last accessed September 2024).
- Kowarik, S., M.-T. Hussels, S. Chruscicki, S. Münzenberger, A. Lämmerhirt, P. Pohl, and M. Schubert (2020). Fiber optic train monitoring with distributed acoustic sensing: Conventional and neural network data analysis, *Sensors* **20**, no. 2, 450, doi: [10.3390/s20020450](https://doi.org/10.3390/s20020450).
- Li, Y., M. Karrenbach, and J. Ajo-Franklin (2022). *Distributed Acoustic Sensing in Geophysics: Methods and Applications*, Vol. 268, John Wiley and Sons, New Jersey.
- Li, Y. E., E. Nilot, Y. Zhao, and G. Fang (2021). Urban activity monitoring using wireless geophones in Singapore, *First International Meeting for Applied Geoscience and Energy Expanded Abstracts*, 3209–3214, doi: [10.1190/segam2021-3583803.1](https://doi.org/10.1190/segam2021-3583803.1).
- Lindsey, N. J., and E. R. Martin (2021). Fiber-optic seismology, *Annu. Rev. Earth Planet. Sci.* **49**, no. 1, 309–336, doi: [10.1146/annurev-earth-072420-065213](https://doi.org/10.1146/annurev-earth-072420-065213).
- Lindsey, N. J., S. Yuan, A. Lellouch, L. Gualtieri, T. Lecocq, and B. Biondi (2020). City-scale dark fiber das measurements of infrastructure use during the covid-19 pandemic, *Geophys. Res. Lett.* **47**, no. 16, e2020GL089931, doi: [10.1029/2020GL089931](https://doi.org/10.1029/2020GL089931).
- Liu, J., S. Yuan, B. Luo, B. Biondi, and H. Y. Noh (2023). Turning telecommunication fiber-optic cables into distributed acoustic sensors for vibration-based bridge health monitoring, *Struct. Contr. Health Monit.* **2023**, 1–14, doi: [10.1155/2023/3902306](https://doi.org/10.1155/2023/3902306).
- Ma, Y., L. Gao, and J. Bi (2023). Three-dimensional high-resolution velocity structure imaging and seismicity study of the tangshan fault, *Geosci. J.* **25**, no. 4, 495–506, doi: [10.1007/s12303-020-0036-5](https://doi.org/10.1007/s12303-020-0036-5).
- Oppenheim, A. V. (1999). *Discrete-Time Signal Processing*, Pearson Education India, Chennai.
- Rashidi, P., D. J. Cook, L. B. Holder, and M. Schmitter-Edgecombe (2011). Discovering activities to recognize and track in a smart environment, *IEEE Trans. Knowl. Data Eng.* **23**, no. 4, 527–539, doi: [10.1109/TKDE.2010.148](https://doi.org/10.1109/TKDE.2010.148).
- Southern California Earthquake Data Center (SCEDC) (2013). *SCEDC Earthquake Data AWS Public Dataset*, Southern California Earthquake Center, Los Angeles, California, doi: [10.7909/C3WD3xH1](https://doi.org/10.7909/C3WD3xH1).
- Shatalin, S. V., V. N. Treschikov, and A. J. Rogers (1998). Interferometric optical time-domain reflectometry for distributed optical-fiber sensing, *Appl. Opt.* **37**, no. 24, 5600–5604.
- Sheng, X., C. Jones, and D. Thompson (2004). A theoretical study on the influence of the track on train-induced ground vibration, *J. Sound Vib.* **272**, no. 3, 909–936, doi: [10.1016/S0022-460X\(03\)00781-8](https://doi.org/10.1016/S0022-460X(03)00781-8).
- Spica, Z. J., J. Ajo-Franklin, G. C. Beroza, B. Biondi, F. Cheng, B. Gaite, B. Luo, E. Martin, J. Shen, C. Thurber et al. (2023). PubDAS: A public distributed acoustic sensing datasets repository for geosciences, *Seismol. Res. Lett.* **94**, no. 2A, 983–998, doi: [10.1785/0220220279](https://doi.org/10.1785/0220220279).
- Talaat, F. M., and H. ZainEldin (2023). An improved fire detection approach based on YOLO-v8 for smart cities, *Neural Comput. Appl.* **35**, no. 28, 20,939–20,954, doi: [10.1007/s00521-023-08809-1](https://doi.org/10.1007/s00521-023-08809-1).
- Ultralytics (2024). Train - Ultralytics YOLO Docs, available at <https://docs.ultralytics.com/modes/train/> (last accessed December 2024).
- Ultralytics (2025). Ultralytics YOLO Docs - YOLO performance metrics, available at <https://docs.ultralytics.com/guides/yolo-performance-metrics/case-studies> (last accessed December 2024).
- Wada, K. (2018). labelme: Image polygonal annotation with python, available at <https://github.com/wkentaro/labelme> (last accessed August 2024).
- Wang, H., Y. Chen, R. Min, and Y. Chen (2022). Urban das data processing and its preliminary application to city traffic monitoring, *Sensors* **22**, no. 24, 9976, doi: [10.3390/s22249976](https://doi.org/10.3390/s22249976).
- Wang, X., Z. Zhan, E. F. Williams, M. G. Herráez, H. F. Martins, and M. Karrenbach (2021). Ground vibrations recorded by fiber-optic cables reveal traffic response to COVID-19 lockdown measures in Pasadena, California, *Commun. Earth Environ.* **2**, 1, 160, doi: [10.1038/s43247-021-00234-3](https://doi.org/10.1038/s43247-021-00234-3).
- Wang, Z., S. Lou, S. Liang, and X. Sheng (2020). Multi-class disturbance events recognition based on EMD and XGBoost in ϕ -otdr, *IEEE Access* **8**, 63,551–63,558, doi: [10.1109/ACCESS.2020.2984022](https://doi.org/10.1109/ACCESS.2020.2984022).
- Williams, E. F., M. R. Fernández-Ruiz, R. Magalhaes, R. Vanthillo, Z. Zhan, M. González-Herráez, and H. F. Martins (2019). Distributed sensing of microseisms and teleseisms with submarine dark fibers, *Nat. Commun.* **10**, no. 1, 5778, doi: [10.1038/s41467-019-13262-7](https://doi.org/10.1038/s41467-019-13262-7).
- Yuan, S., M. van den Ende, J. Liu, H. Y. Noh, R. Clapp, C. Richard, and B. Biondi (2024). Spatial deep deconvolution u-net for traffic analyses with distributed acoustic sensing, *IEEE Trans. Intell. Transp. Syst.* **25**, no. 2, 1913–1924, doi: [10.1109/TITS.2023.3322355](https://doi.org/10.1109/TITS.2023.3322355).
- Zhan, Z. (2019). Distributed acoustic sensing turns fiber-optic cables into sensitive seismic antennas, *Seismol. Res. Lett.* **91**, no. 1, 1–15, doi: [10.1785/0220190112](https://doi.org/10.1785/0220190112).
- Zhang, G., Z. Song, A. G. Osotuyi, R. Lin, and B. Chi (2022). Railway traffic monitoring with trackside fiber-optic cable by distributed acoustic sensing technology, *Front. Earth Sci.* **10**, 990837, doi: [10.3389/feart.2022.990837](https://doi.org/10.3389/feart.2022.990837).
- Zhu, T., J. Shen, and E. R. Martin (2021). Sensing earth and environment dynamics by telecommunication fiber-optic sensors: An urban experiment in Pennsylvania, USA, *Solid Earth* **12**, no. 1, 219–235, doi: [10.5194/se-12-219-2021](https://doi.org/10.5194/se-12-219-2021).
- Zhu, W., E. Biondi, J. Li, J. Yin, Z. E. Ross, and Z. Zhan (2023). Seismic arrival-time picking on distributed acoustic sensing data using semi-supervised learning, *Nat. Commun.* **14**, no. 1, 8192, doi: [10.1038/s41467-023-43355-3](https://doi.org/10.1038/s41467-023-43355-3).

Appendix

Label transfer data processing length

The duration of the data is related to the cost of data annotation. At the same downsampling rate (e.g., downsampled to 10 Hz), visualizing and manually annotating a 4-min data sample (as shown in Fig. A1) results in 93 training samples obtained through label transfer (with a timestep of 2.56 s). In comparison, an 8-min data sample can produce 187 training samples with a

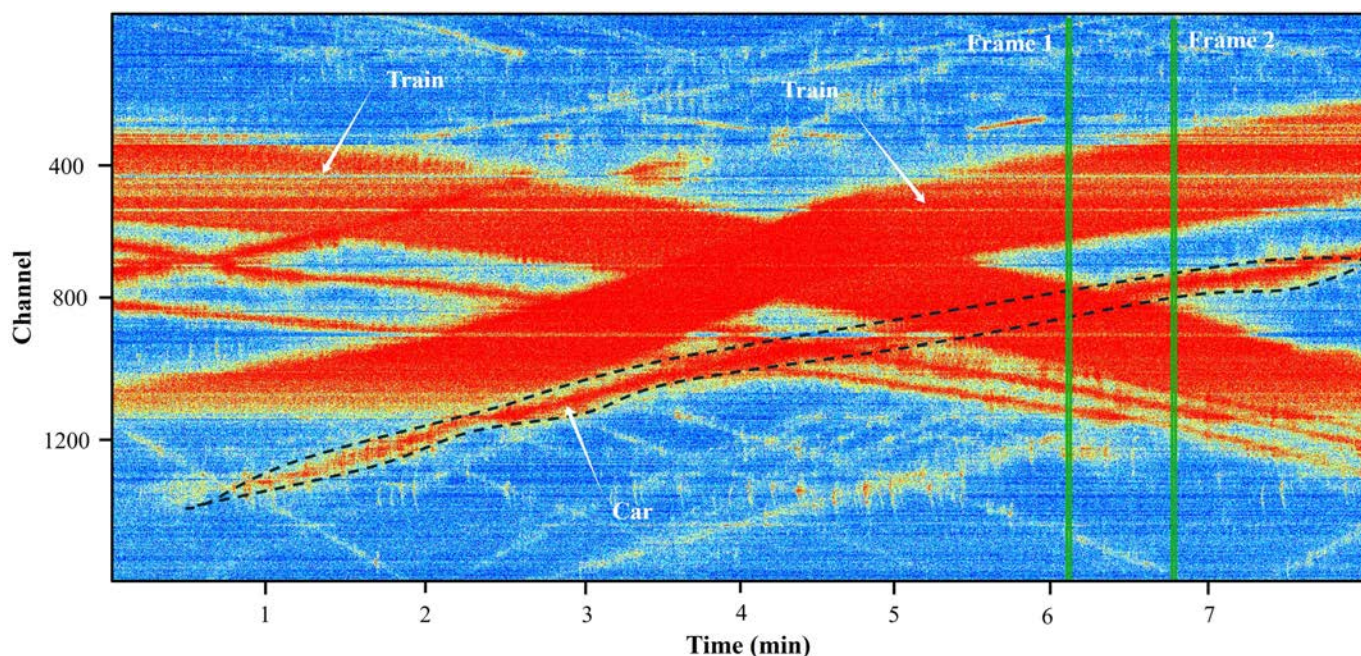


similar level of manual annotation cost. However, the data duration cannot be increased indefinitely because this would make the visualized images excessively long and significantly increase the difficulty of annotation. For instance, an 8-min data sample downsampled to 10 Hz yields 4800 points. Considering our data has 1580 channels, the resulting image size for manual annotation is 4800×1580 , as shown in Figure A1, in which the width is three times the height. This dataset is characterized by frequent traffic signals, with most annotations being polygons representing vehicle trajectories, so wider images are easier to annotate. By visualizing data samples of different lengths, we found that this image size balances ease of manual annotation with maximizing sample conversion rate (i.e., one spatiotemporal sample can be converted into 187 space–frequency domain samples through label transfer). Therefore, an 8-min duration was chosen for spatiotemporal visualization and annotation of this dataset.

Figure A1. A comparison chart of different data lengths in label transfer. The color version of this figure is available only in the electronic edition.

Handling overlapping signal sources

When labeling the data using the label transfer method, we sometimes encounter situations where two signal sources overlap. Object overlap is a common challenge in computer vision tasks. In such cases, we will take the following measures to address it. First, we aim to avoid overlap during the selection of training data for annotation. During the annotation process, when encountering overlapping points, our usual approach is that the signal source with stronger energy will override the signal source with weaker energy. As shown in Figure A2, which provides a time–space visualization of a section of



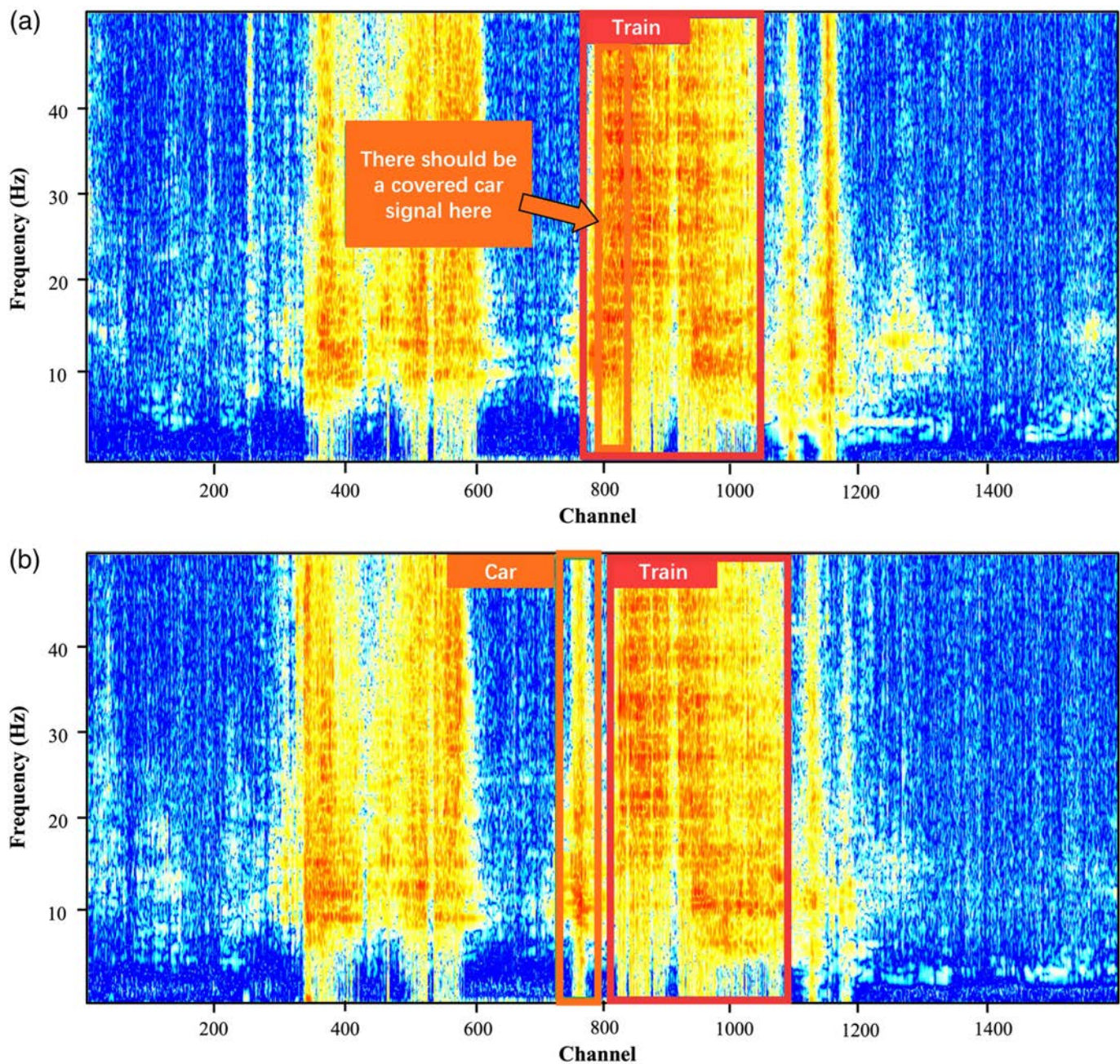
distributed acoustic sensing (DAS) data, the vehicle trace, highlighted by the dashed line, overlaps with a train trace. To further analyze this, we extracted data from two specific timeframes and generated their corresponding space–frequency spectrograms, with frame 1 and frame 2 displayed in Figure A3a,b. As illustrated in Figure A3, the vibration energy generated by the train is significantly greater than that of the car. Consequently, at the point where the train and car signals overlap, it becomes challenging to manually distinguish the vehicle signals from the stronger train signals. When the train signal and car signal overlap, we typically label the overlapping part as the train in the time–space domain. As shown in Figure A4, the red (train) label covers the yellow (car) label.

On the other hand, to enhance data diversity, we also manually labeled some overlapping samples, especially when the energy levels of the two signal sources are not significantly different. As shown in Figure A5, the train and seismic signal sources have overlapping parts on the spectrogram. We labeled each object on the spectrogram separately for model training, which enhances the model’s generalization ability. This enhances the model’s ability to recognize overlapping acoustic sources. As shown in Figure A6, the signal strengths of factory and vehicle sources are comparable, and the model can differentiate between the two overlapping source signals. However, in cases where the signals from trains and cars overlap, as illustrated in Figure A3, the signal strength of the train is much greater than that of the car, making it difficult for humans to distinguish the car source covered by the train, and the model performs poorly.

Figure A2. Visualizing the overlapping data between train and car. The dashed line highlights a trace of a car. Transform the frame 1 and frame 2 data into the space–frequency domain, resulting in Figure A3a,b, respectively. The color version of this figure is available only in the electronic edition.

Power spectral density analysis of car events

The frequency range for each type of source was not determined arbitrarily but was assigned based on an analysis of the source’s spectrum. Taking vehicle events as an example, as shown in the figures, we randomly selected 12 vehicle source DAS data sets for analysis and plotted the corresponding power spectral density (PSD) diagrams, as shown in Figure A7. We also plotted comparative diagrams of the PSD for these vehicle data, as shown in Figure A8. The car induced vibration components above 50 Hz can be disregarded. On the other hand, the ground vibrations from vehicles are indeed influenced by many factors. As illustrated in Figure A7, waveform recordings of different vehicle signals vary. However, the comparative PSD diagrams of vehicle signals show that the vibration frequency range produced by almost all conventional vehicles is primarily concentrated between 5 and 40 Hz, which allows for the determination of the frequency range of the space–frequency box. To capture more detailed frequency characteristics, the frequency observation range for vehicle signals in this study is set to 0–50 Hz. The main goal of this article is to propose a classification method for acoustic signal detection based on urban DAS data. Of course, vehicle-induced ground vibrations influenced by factors such as different speeds and weather



conditions will be analyzed in more detail in our subsequent studies.

Performance on the whole dataset

After testing the proposed method on the entire DAS dataset, we found that it generally performs well in identifying various targets such as trains, cars, and factory production through the entire dataset. In Figure A9, we present several randomly selected spectrogram frames with the predictions to demonstrate the model's performance. However, there are some instances where targets are missed, as shown in the image in the lower-right corner. The signal on the right should be the excitation generated by the source vehicle, but the model's prediction missed this target. This issue may be due to the

Figure A3. Visualization of overlapping train and car signals in the space–frequency domain. (a) Train signals override car signals. (b) Train signals are separated from the car signals. The color version of this figure is available only in the electronic edition.

insufficient number of source vehicle instances in the training dataset. Future improvements could be achieved by expanding the training dataset.

In addition, we input the entire dataset into the model sequentially according to the time series, and arranged the predicted spectrogram results frame by frame in

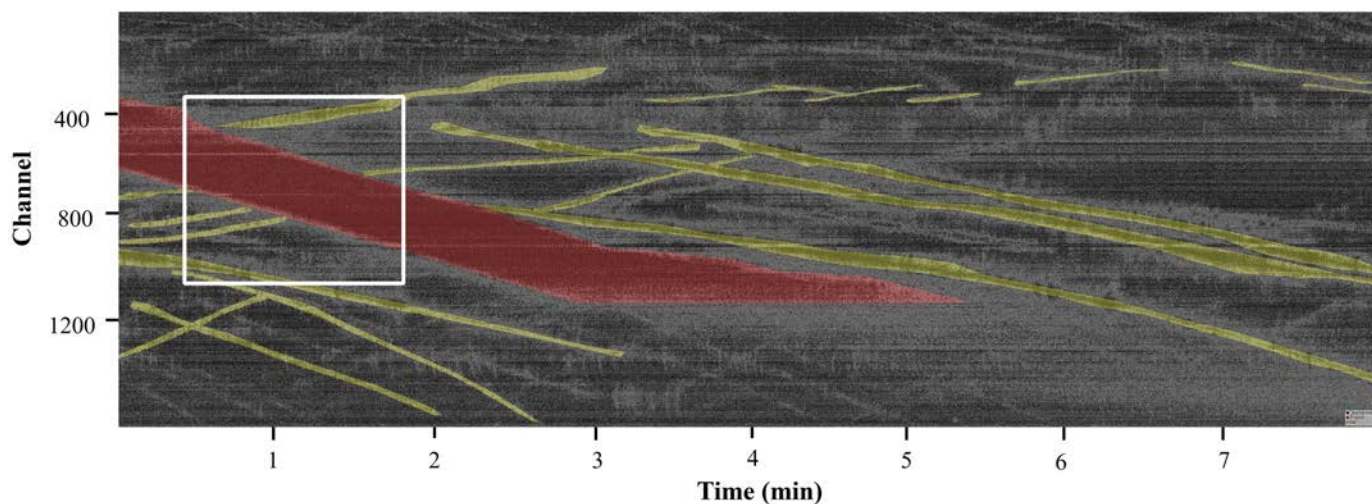


Figure A4. In the manually labeled examples in the spatiotemporal domain, when overlapping categories are encountered, we assume that the signal with stronger energy will override the

signal with weaker energy. The color version of this figure is available only in the electronic edition.

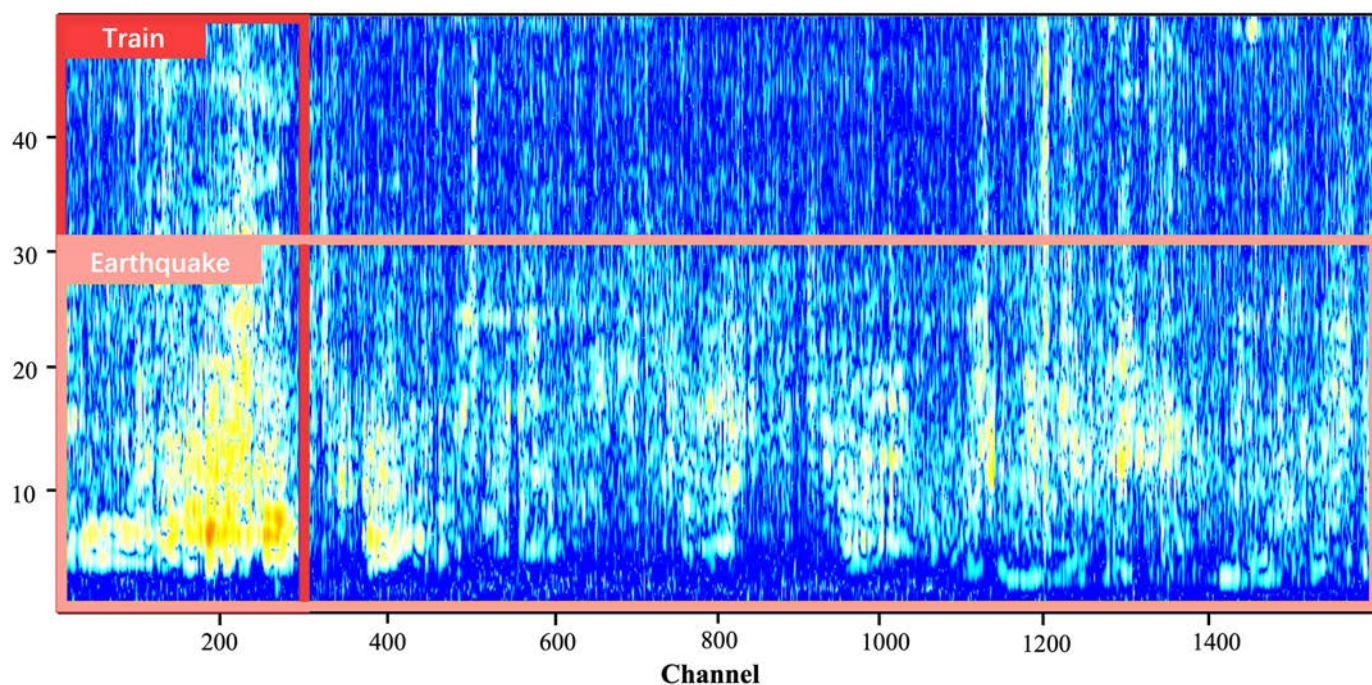
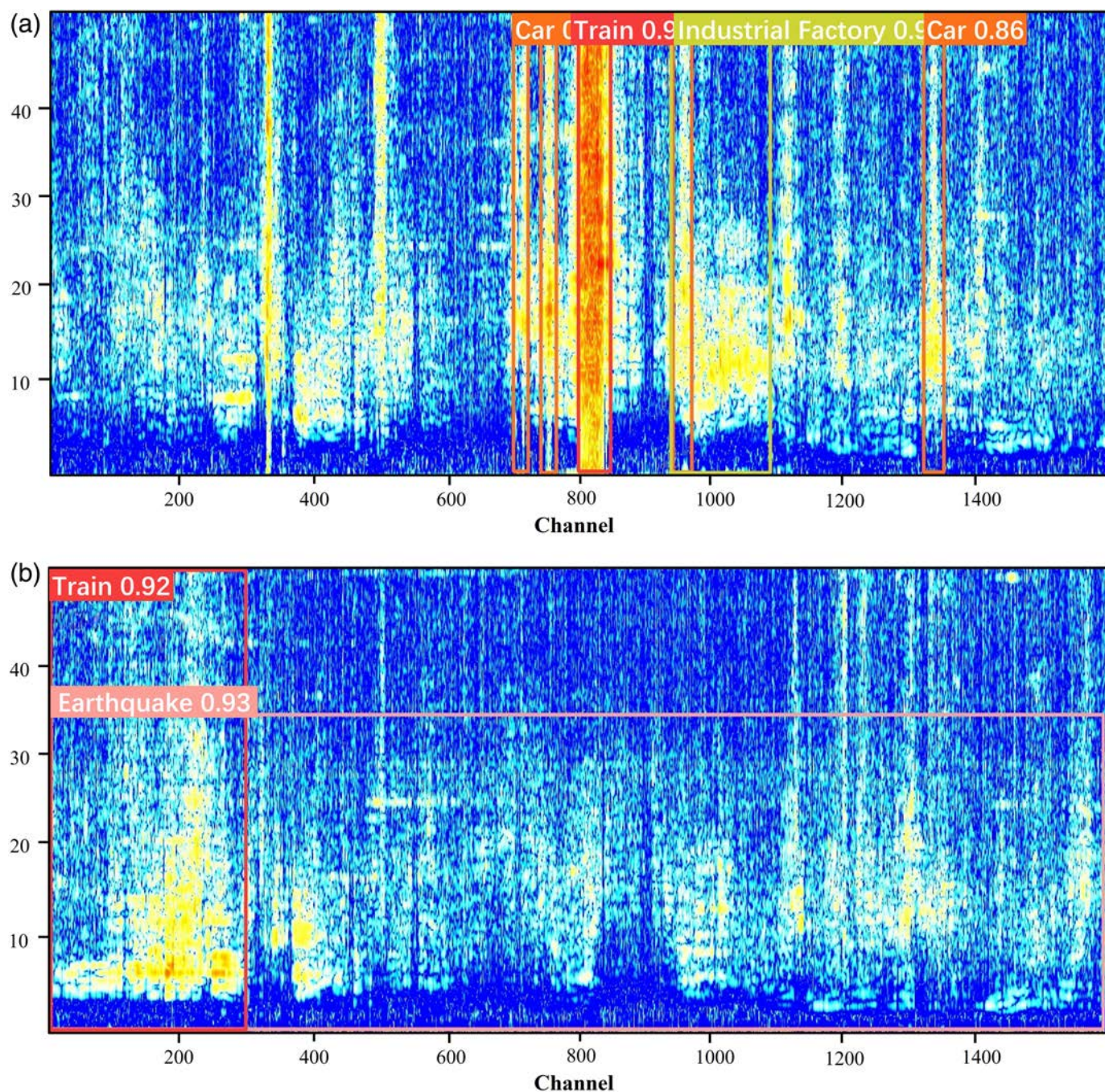


Figure A5. Manual labeling of overlapping tags. Earthquake and train signals are not significantly different, so manually adding

overlapping tag samples can be beneficial. The color version of this figure is available only in the electronic edition.

chronological order, converting them into time-space-domain prediction results, as shown in Figure A10. The time-space domain result can help us more comprehensively understand the model's performance in the whole dataset. As shown in Figure A10, the x -axis represents time, and the y -axis represents the number of channels. We use red lines to

denote predictions for trains, black lines for cars, green lines for factories, orange lines for vibrator vehicles, and blue lines for earthquake events. It can be seen that our model performs exceptionally well across the entire dataset, with clear vehicle trajectories and accurate prediction of earthquake occurrences. We also did compare the train recognition results with

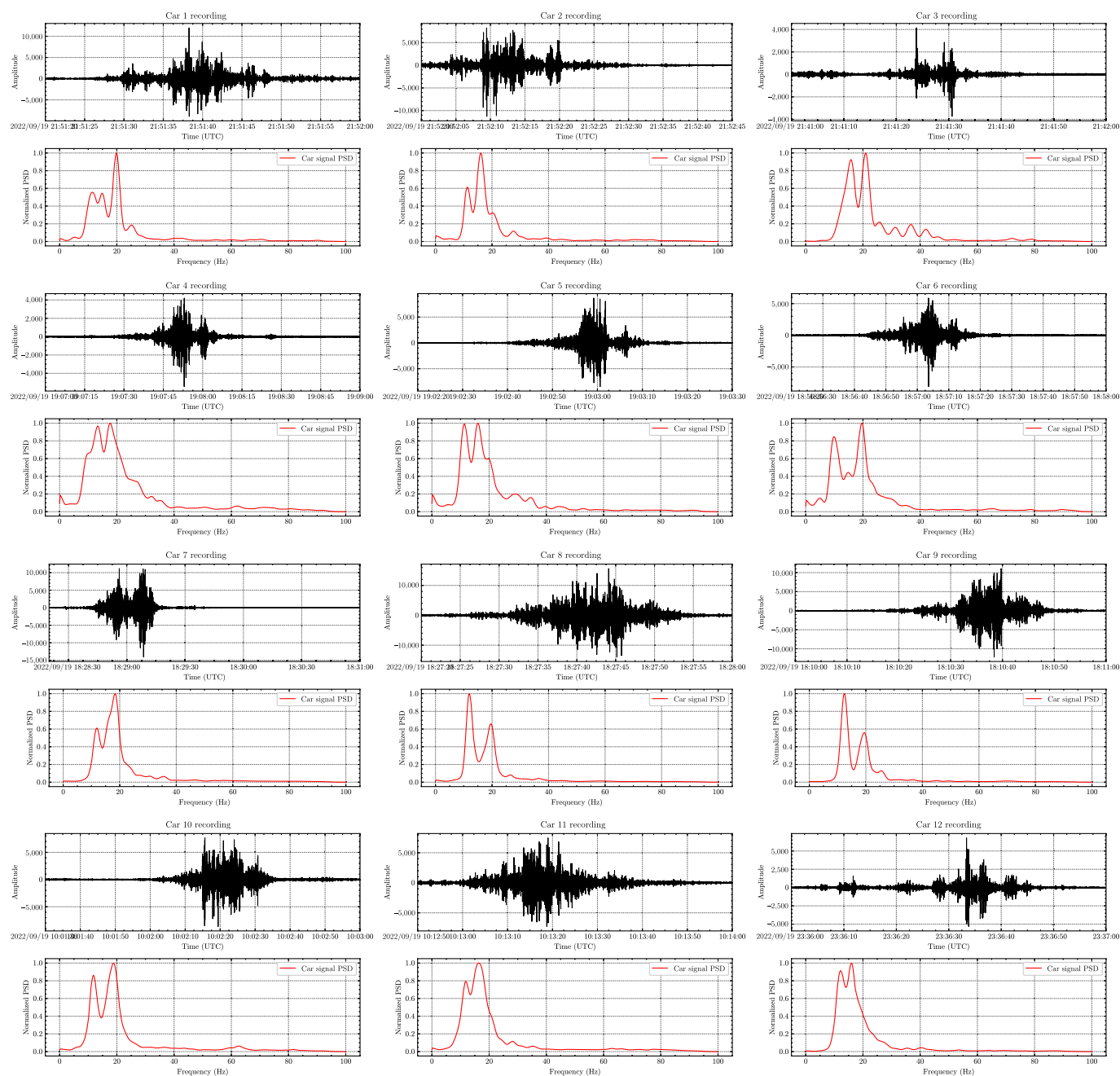


the published passenger train schedules. We found that, in addition to the published train numbers, the recognition results also included some operational trajectories that have not yet been publicly disclosed. These undisclosed trajectories are likely freight train operations because this railway line is one of the main routes for coal transportation from Tangshan, and it should include both regular passenger trains and freight trains. A detailed analysis of the regional traffic recognition results will be conducted in the next phase of our research.

Figure A6. Prediction results for overlap acoustic sources. (a) The industrial factory (green label) and car (orange label) signals have some overlap. (b) The train (red label) and earthquake (pink label) signals have some overlap. The color version of this figure is available only in the electronic edition.

Model generalization analysis

To evaluate the generalization ability of the model proposed in this study, we conducted tests on multiple datasets, including



the foresee dataset published on PubDAS (Spica *et al.*, 2023) and the Southern California Earthquake Data Center (SCEDC) earthquake data AWS Public dataset (SCEDC, 2013) (see [Data and Resources](#)) (Zhu *et al.*, 2023). We detected two types of acoustic events: vehicle and seismic events. Specifically, the FORESEE dataset contains acoustic events for both vehicle and earthquake, whereas the SCEDC dataset focuses on earthquake event detection. The results is shown as follows: Figure A11 showing the detection results for vehicle events, and Figure A12 for earthquake events. For the vehicle detection, as shown in Figure A11, our model was still able to effectively identify the targets. It can be observed that the model accurately locates and identifies vehicles. This robustness

Figure A7. DAS recordings 12 cars and their corresponding power spectral density. The color version of this figure is available only in the electronic edition.

ensures that the model can adapt to and perform well on a variety of data, indicating its versatility and potential for broad application in diverse scenarios.

Figure A12a presents the detection result for a microseismic event (M 1.1) in the FORESEE dataset (Zhu *et al.*, 2021). Despite the small magnitude of this event, our model successfully identified it as “earthquake” class and the associated

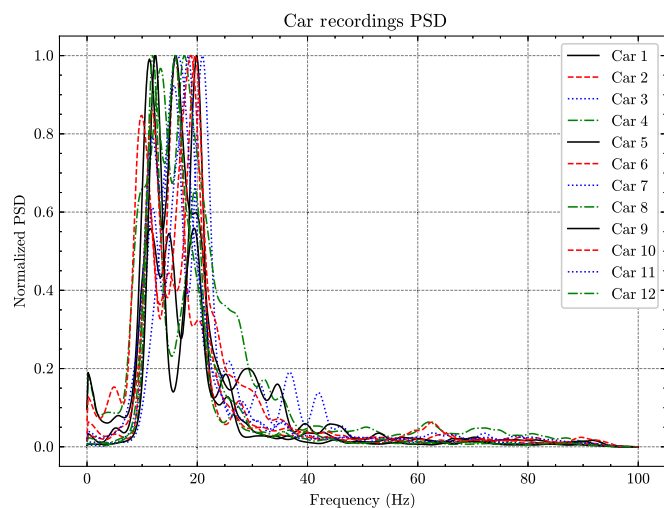


Figure A8. Comparison of power spectral density across multiple cars recordings. The color version of this figure is available only in the electronic edition.

confidence scores of 0.96, marked as “Earthquake 0.96” in Figure A12a, highlighting the method’s sensitivity and reliability in detecting low-magnitude microearthquakes. In the SCEDC earthquake data AWS public dataset, as shown in Figure A12b, our method achieved significantly high confidence scores of 0.97 for an earthquake event in that dataset, demonstrating excellent detection performance. This further validates the model’s robustness across earthquake events with varying frequency and apparent velocity characteristics.

The experimental results further confirm our model’s effectiveness and generalization ability in detecting various acoustic events. However, it is important to note that if our model is used as a pretrained model and subsequently fine-tuned on a specific dataset, better results could be achieved. Fine-tuning the model on domain-specific data would allow it to better adapt to the unique characteristics of that dataset, potentially leading to improved performance.

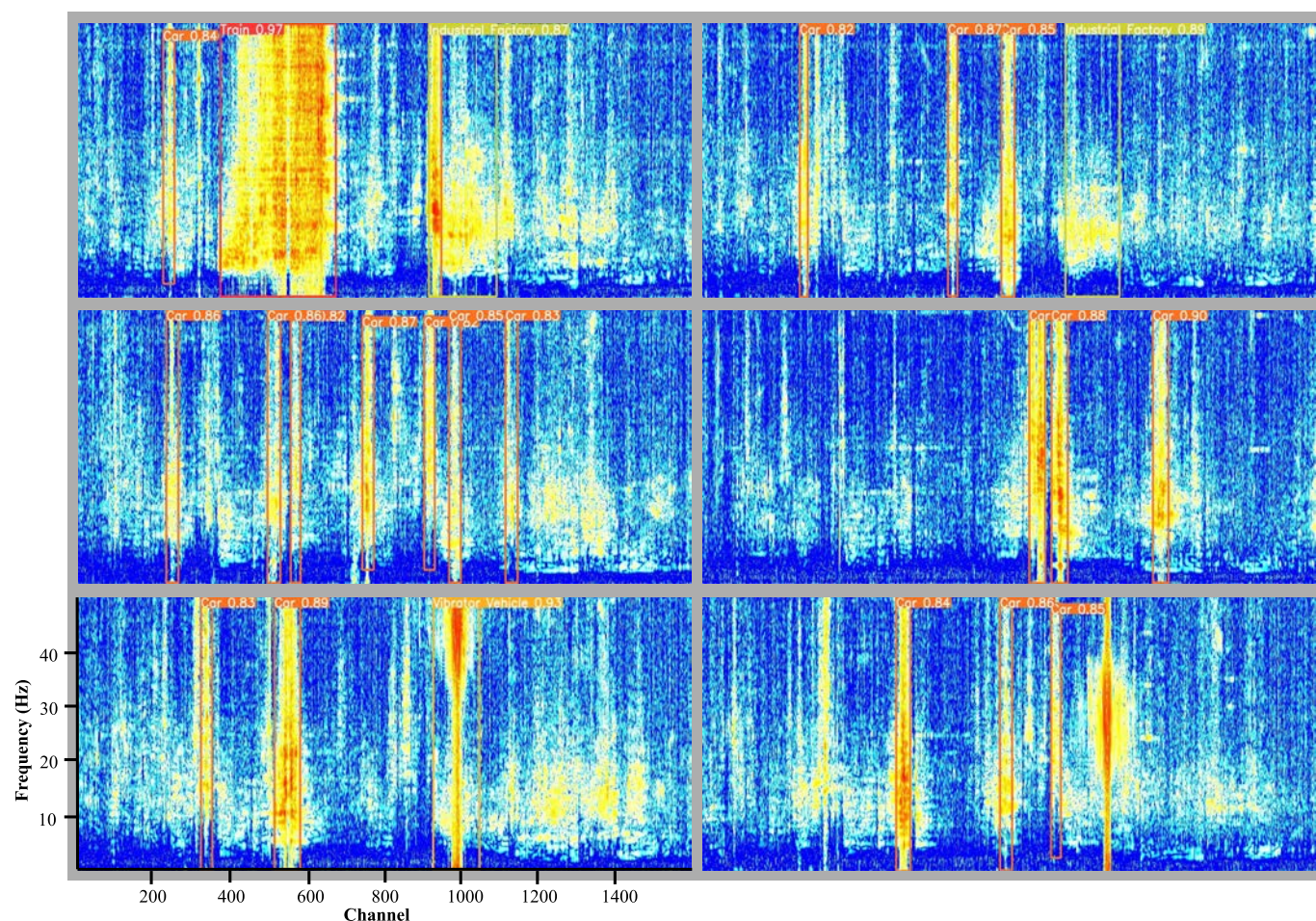


Figure A9. Randomly selected spectrogram frames with predictions from the dataset. Each subplot showcases predictions of different events within the dataset. The color coding for event types is consistent with the main text: red indicates trains, orange indicates

cars, yellow indicates vibrator vehicles, pink represents earthquakes, and green denotes industrial factory signals. The x-axis represents the channel, and the y-axis represents the frequency (Hertz). The color version of this figure is available only in the electronic edition.

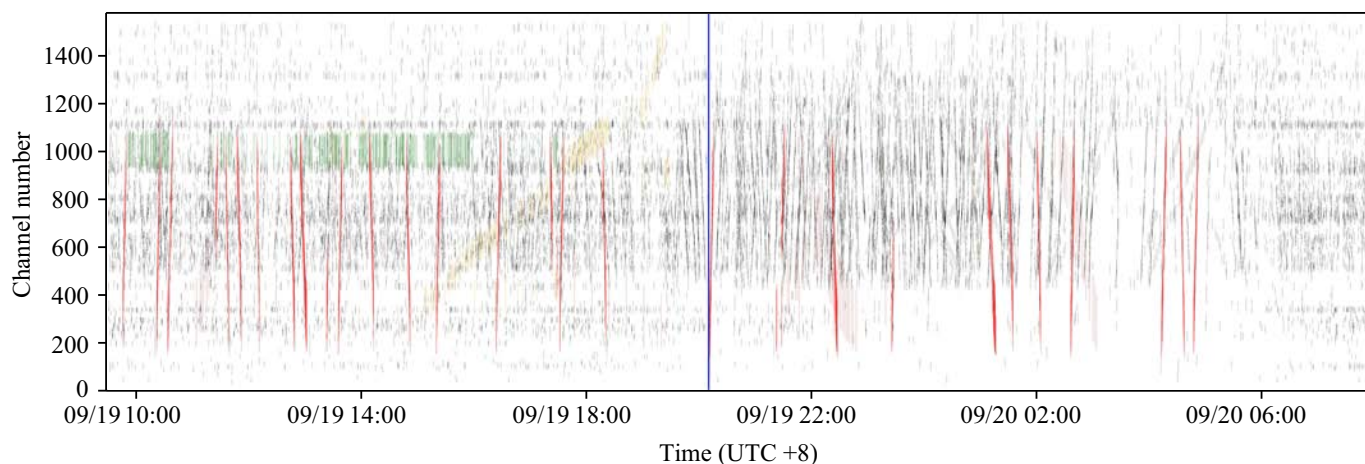


Figure A10. Time–space domain representation of total dataset predictions. The x-axis represents time, and the y-axis represents channel number. Colored lines indicate different detected events: red for trains, black for cars, green for factories, yellow for vibrator vehicles, and blue for earthquake events. This

comprehensive visualization provides an overview of the spatial and temporal distribution of acoustic events detected by the model across the entire dataset. The color version of this figure is available only in the electronic edition.

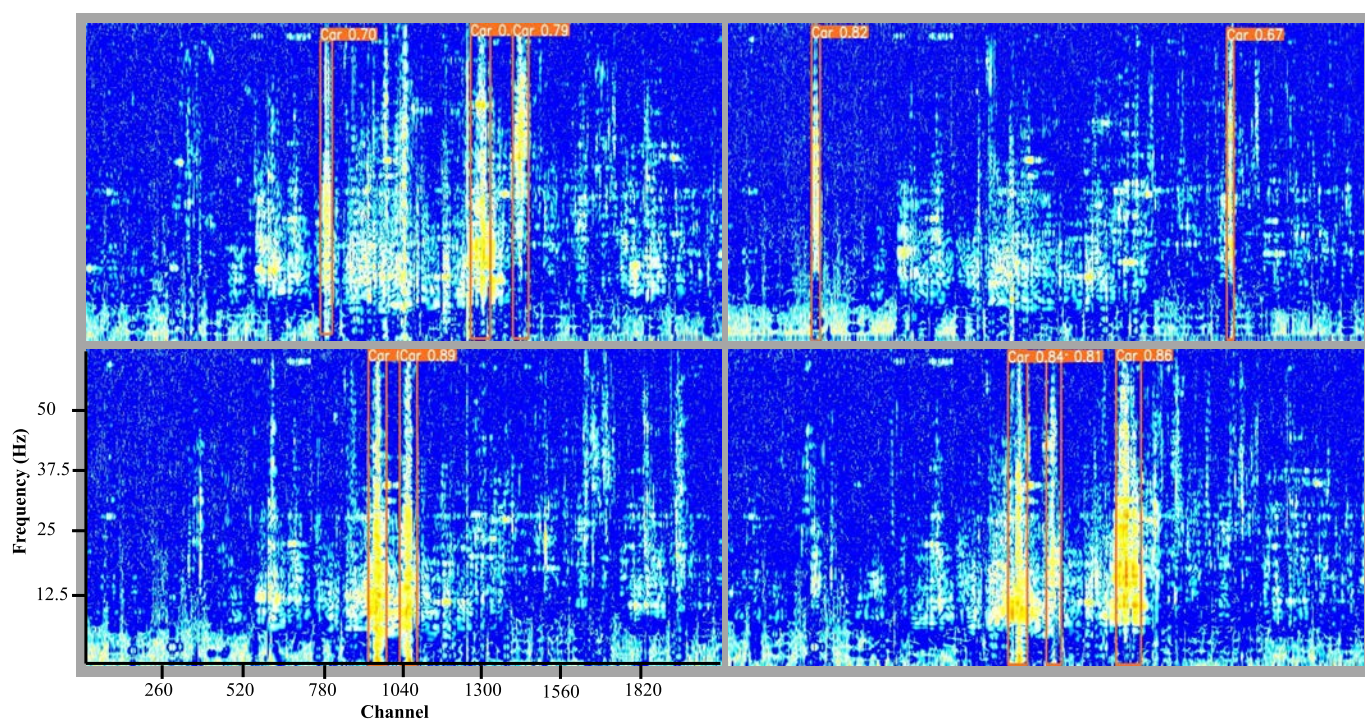


Figure A11. Detection of vehicle events in the FORESEE DAS dataset. For each subplot, the orange unfilled rectangular boxes highlight the detected signals corresponding to vehicle events.

The x-axis represents the channel, and the y-axis represents the frequency (Hertz). The color version of this figure is available only in the electronic edition.

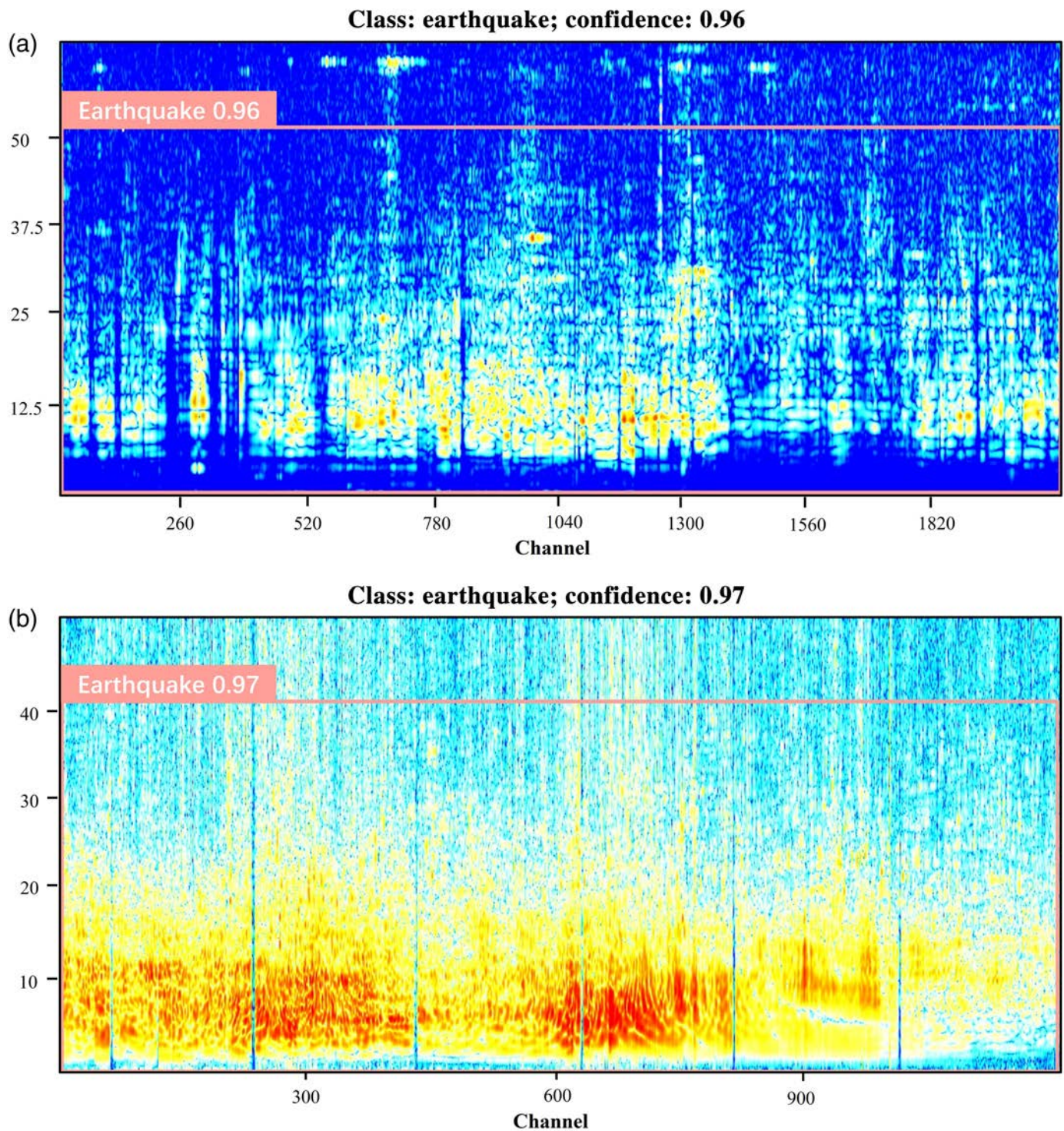


Figure A12. Earthquake detection results using DAS data from the FORESEE and Southern California Earthquake Data Center (SCEDC) datasets. The x-axis represents the DAS channel number, and the y-axis indicates the frequency range in Hertz. The pink unfilled rectangular boxes highlight the signals corresponding to the detected earthquake events. The labels

“Earthquake 0.96” and “Earthquake 0.97” within the filled pink boxes denote the detected earthquake class and the associated confidence scores. (a) Detection results for a single earthquake event from the FORESEE dataset. (b) Detection results for a single earthquake event from the SCEDC dataset. The color version of this figure is available only in the electronic edition.