西安欧亚学院 eurasia UNIVERSITY

Automatic Detection of Seismic and Infrasonic Signals from Nuclear Explosions Based on Machine Vision (O2.2-116)

Bridging the Paradigm Gap Between Manual Analysis and Automated Processing

CHEN Huhu

Xi'an Eurasia University

CTBT: Science and Technology 2025 Conference

Vienna, Austria, 8-12 September 2025

Session starts at 16:00

Agenda

1. Introduction & The Core Problem: The "Paradigm Gap"
2. Proposed Innovation: A Shift from Signal Processing to Visual Perception
3. Methodology: A 4-Step Machine Vision Framework
    - 3.1. Signal Preprocessing & Imagification
    - 3.2. Dataset Construction & Annotation
    - 3.3. Model Training with YOLOv11
    - 3.4. Inference & Application
4. Experimental Results & Analysis
    - 4.1.Experimental Setup
    - 4.2 Quantitative Performance
5. Conclusion & Future Work

1 Introduction

1.1 The Critical Role of CTBT Monitoring

The Comprehensive Nuclear-Test-Ban Treaty (CTBT) relies on the International Monitoring System (IMS) to detect unauthorized nuclear tests, with four key monitoring technologies: seismic, infrasound, hydroacoustic, and radionuclide. My research focuses specifically on seismic and infrasonic signals—these generate massive streams of continuous waveform data, but the grand challenge remains: automatically, rapidly, and accurately distinguishing nuclear explosion signals from natural events (e.g., earthquakes, storms) and background noise.

1 Introduction

1.2The Core Problem: The "Paradigm Gap"

At the heart of this challenge lies a “paradigm gap”:
Current state-of-the-art automated systems: Depend on 1-D signal processing techniques—such as STA/LTA or slowness-based triggers, wavelet transforms or hand-crafted feature extraction (e.g., duration, amplitude, spectral ratios), and traditional ML classifiers like SVM or Random Forest. These methods struggle to capture the intuitive, context-driven patterns that human experts recognize.

Expert human analysts: Rely on visual interpretation. They examine waveform plots and spectrograms, identifying subtle shapes, temporal patterns, and contextual cues that are instinctive to the eye but extremely difficult to mathematically encode into 1-D algorithms.

This gap leads to a critical limitation: automated results still fall short of manual review. Final human validation remains mandatory, which is inefficient, time-consuming, and unscalable for the IMS’s ever-growing data volume. So, how do we bridge this divide?

2: Our Innovative Proposal: A Paradigm Shift

Our solution aligns with how experts actually work: if human analysts use vision to identify signals, we should teach AI to "see" too. The core idea is to reframe the 1D signal detection problem as a 2D visual recognition task—a paradigm shift that lets AI learn the same visual patterns and features that experts rely on.

Here's how it works in practice: we convert 1-D time-series seismic and infrasonic signals into 2-D images (e.g., waveform plots, spectrograms), then apply state-of-the-art computer vision models to these images. By doing so, we effectively encode human visual expertise into the automated system—closing the gap between manual intuition and machine processing.

3: Methodology: Overall 4-Step Framework

To turn this idea into a actionable pipeline, we designed a cl
4-step framework. Let’ s break down each step:

Input

Raw Seismic/Infrasound Signal

Step 1: Preprocessing & Imagification

Signal -> Image (Spectrogram/Waveform Plot)

Step 2: Dataset Construction: Image -> Manually

Annotated Image (Bounding Boxes for P, S, I waves)

Step 3: Model Training

Annotated Images + YOLOv11 Algorithm -> Trained Model

Step 4: Inference & Application:

New Signal -> Image -> Trained Model -> Output (Class + Time Bounding Box,signal type, P/S label)

Step 1: Signal Preprocessing & Imagification

First, we refine raw signals and transform them into high-quality, feature-rich images—this is the foundation of our visual approach:

Preprocessing: We address noise and inconsistencies with four key steps:

- **Eliminate breakpoint abnormal signals to remove artifacts.**
- **Detrend to strip out long-term instrumental or environmental trends (e.g., drift from sensor temperature changes).**
- **Apply bandpass filtering to isolate relevant frequency bands (e.g., 1–10 Hz for P-waves, 0.3–4 Hz for infrasound).**
- **Normalize amplitude to a standard range (e.g., [0, 1024]) to ensure consistent input for the model.**

Imagification: We convert processed signals into two types of 2-D images, each serving a purpose:

- **Waveform plots: Simple line graphs of amplitude vs. time—ideal for clear, impulsive signal phases (e.g., sharp P-waves).**
- **Spectrograms: Time-frequency representations where color intensity indicates power at a specific frequency and time. These are far richer in features and superior for distinguishing complex signals (e.g., overlapping S-waves or infrasound I-waves).**

Step 2: Dataset Construction & Annotation

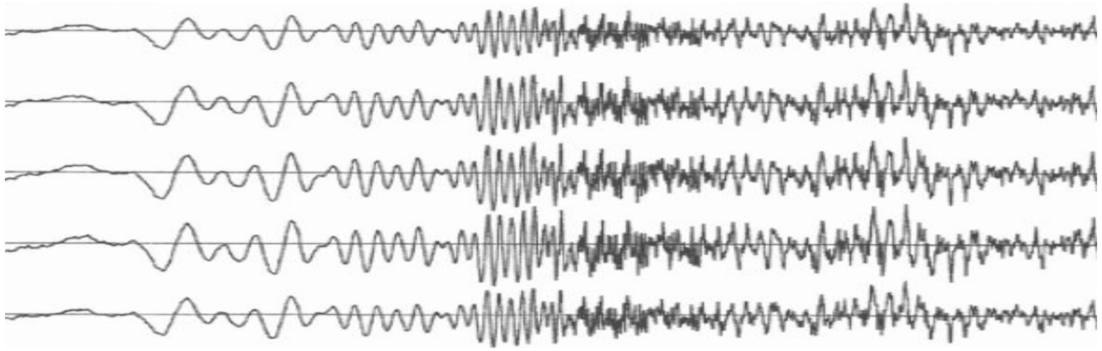
Dataset Construction

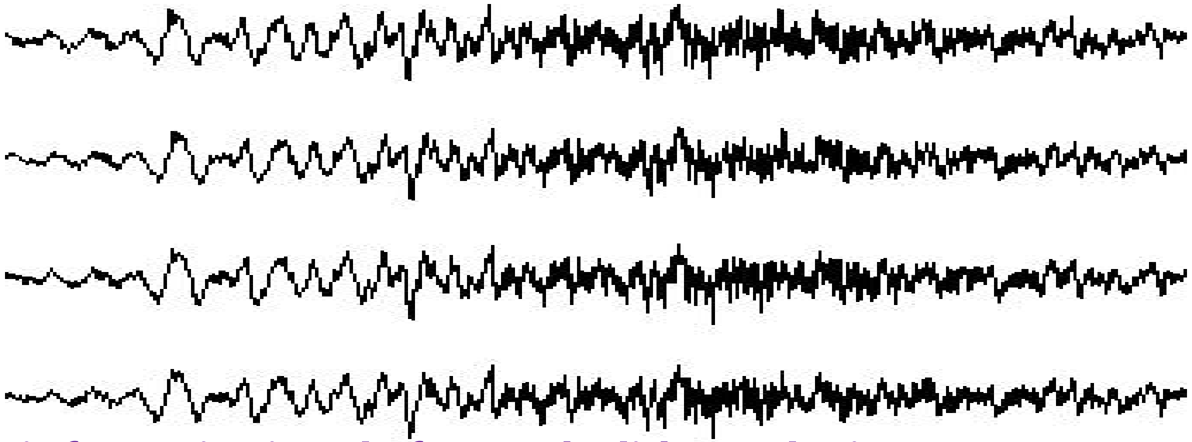
A robust computer vision model needs high-quality labeled data—so we built a curated dataset tailored to nuclear explosion monitoring:

Data curation: We collected confirmed "ground-truth" data, including undergroud seismic signals from nuclear and mining explosions, infrasonic signals from historical atmospheric nuclear explosion and bolides explosion signals. We also included background noise (e.g., microbaroms, MAWs) to ensure the model learns to distinguish signals from interference.

Data examples of atmospheric nuclear explosion infrasound

infrasonic signals from one of the atmospheric nuclear explosions from China

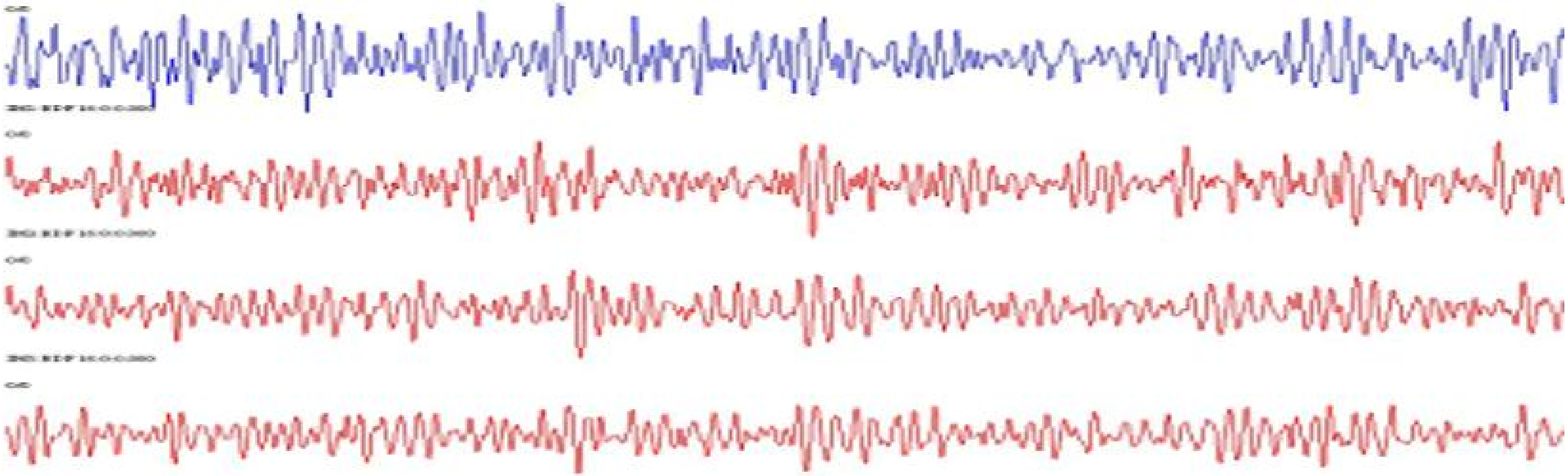


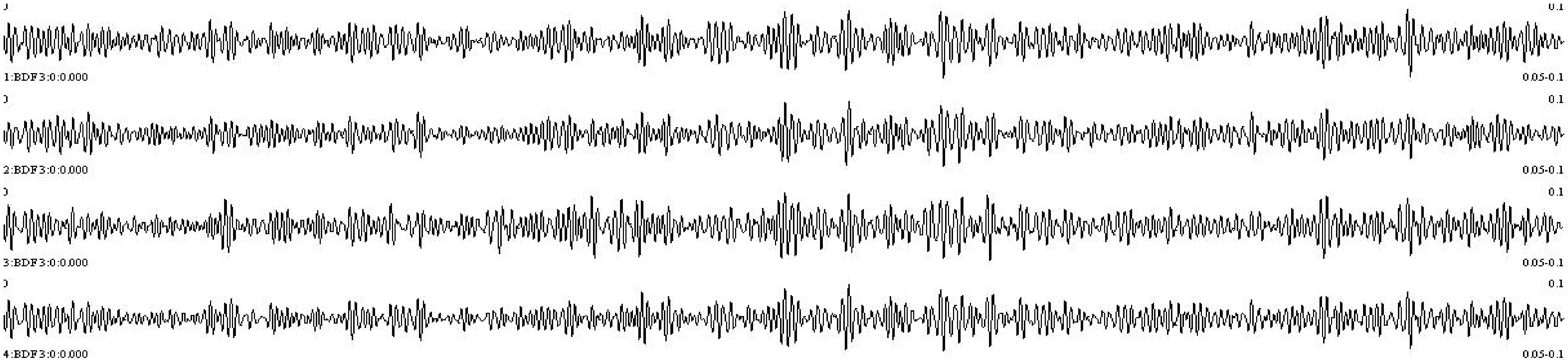


infrasonic signals from a bolide explosion

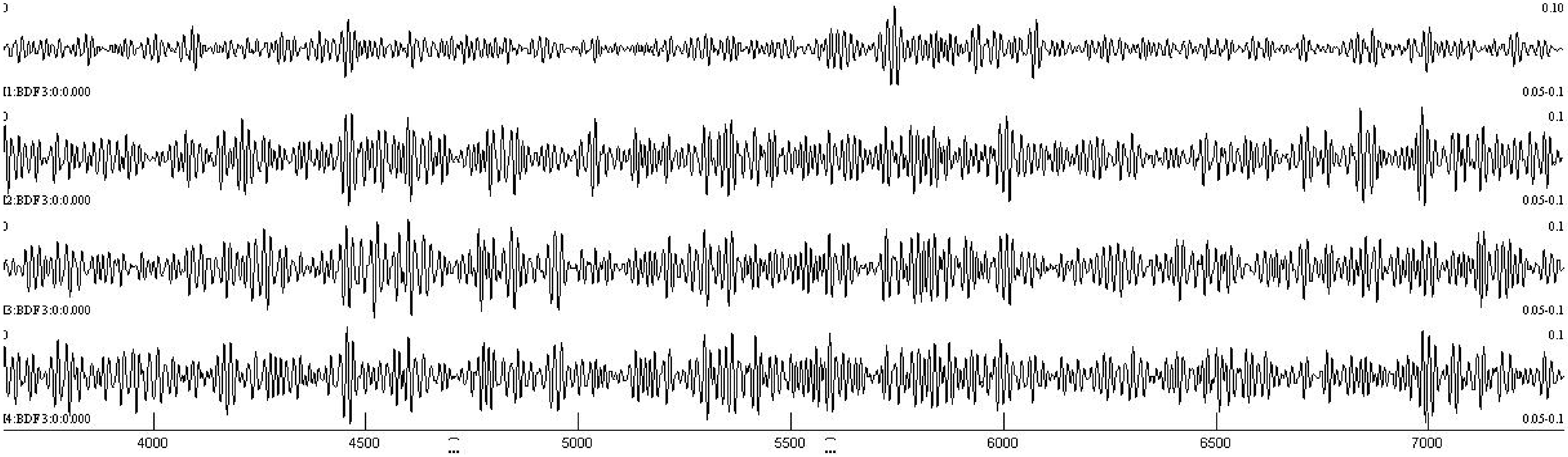
0.02-5Hz

record noises of MAWs and microbaroms of long durations continuously almost every day.





0.1
1:BDF3:0:0.000
0.05-0.1
0.1
2:BDF3:0:0.000
0.05-0.1
0.1
3:BDF3:0:0.000
0.05-0.1
0.1
4:BDF3:0:0.000
0.05-0.1



0.10
1:BDF3:0:0.000
0.05-0.1
0.1
2:BDF3:0:0.000
0.05-0.1
0.1
3:BDF3:0:0.000
0.05-0.1
0.1
4:BDF3:0:0.000
0.05-0.1
4000
4500
5000
5500
6000
6500
7000

MAWs

Step 2: Dataset Construction & Annotation

Annotation:

Using tools like LabelImg, we manually drew bounding boxes around key signals. Our annotation scheme is precise and task-specific:

- For seismic signals: Label signal classes (explosion, earthquake) and seismic phases—specifically P-waves (primary) and S-waves (secondary)—with the flexibility to extend to sub-phases like Pn, Pg, Sn, or Lg. Notably, an annotated image typically contains multiple labels, covering both signal class and seismic phases.
- For infrasound signals: Label signal classes (explosion) and infrasound phases—namely I-waves—including sub-types like Ig (Ground), Is (Stratosphere), or It (Thermosphere). Similarly, an annotated image usually includes multiple labels, encompassing both signal class and infrasound phases (e.g., "explosion", "Is").

Critically, each bounding box defines two things: the signal’s class (e.g., “explosion”) and its exact time window (start/end time)—this is essential for later event localization.

Step 3: AI Model select and Training

- Why YOLOv11?

We chose YOLOv11 (You Only Look Once) as our core model—and for good reason: it’s the latest evolution of real-time object detection architectures, and it’s perfectly suited to our task.

Why YOLOv11?

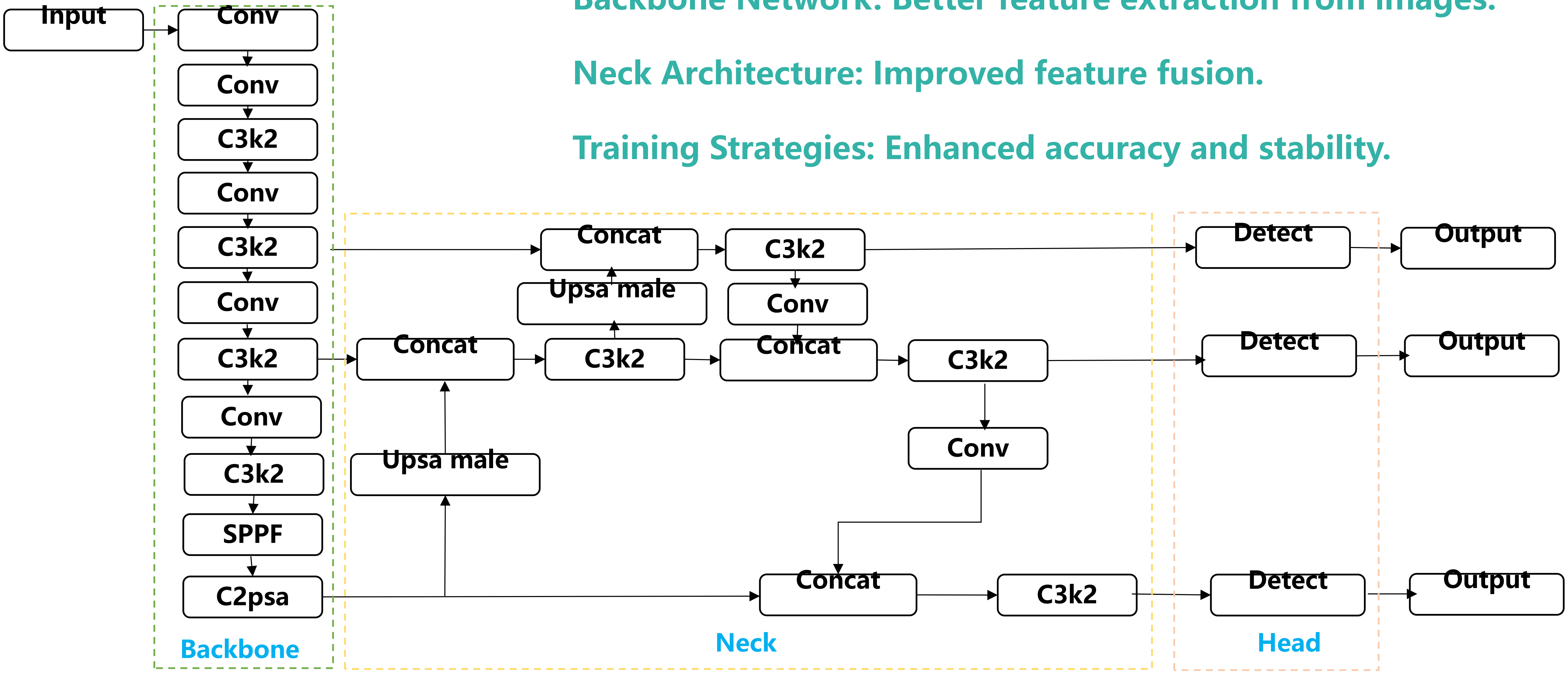
- Single-stage detector: Extremely fast and efficient—critical for processing the IMS’s continuous data streams in near-real time.
- High accuracy: It features improved components:
- Backbone network: Better at extracting fine-grained features from signal images (e.g., subtle frequency shifts in spectrograms).
- Neck architecture: Enhances feature fusion, helping the model combine low-level (e.g., edge) and high-level (e.g., pattern) features.
- Training strategies: Boosts accuracy and stability, even with limited labeled data.
- Multi-object detection: It can identify multiple signals (e.g., P-wave + S-wave) in a single image—unlike some 1-D methods that struggle with overlapping phases.

Step 3: AI Model select and Training - Why YOLOv11?

Backbone Network: Better feature extraction from images.

Neck Architecture: Improved feature fusion.

Training Strategies: Enhanced accuracy and stability.



Step 3: AI Model select and Training

How to train the models?

Our training setup was straightforward but rigorous:
We used the YOLO framework (via Ultralytics) with the pre-trained yolo11n.pt model as a starting point.

Dataset split: 50% training, 10% validation, 40% testing—ensuring we evaluate generalization, not just overfitting.
Training parameters: 200 epochs, input image size of 1024x2048 , batch size of 16.

Classes labeled: 5 key categories— ‘p’ (P-wave), ‘s’ (S-wave), ‘explosion’ , ‘earthquake’ ,’ I’ (infrasonic wave),

Step 4: Inference & Application

Deployment: The trained YOLOv11 model is integrated into the processing pipeline.

Process for New Data:

Continuous data is segmented.

Each segment is preprocessed and converted to an image (same as training).

The image is fed into the model.

The model predicts bounding boxes with class probabilities.

Output:

Detection: Whether a signal (P, S, or I) is present.

Classification: What type of signal/event it is (e.g., “explosion” vs. “earthquake”).

Localization: Exact start and end times of the signal—this provides the arrival time data needed to calculate the event’s geographic location (a core requirement for CTBT verification).

4. Experimental Results & Analysis

4.1 Experimental Setup

Dataset:
1,000 seismic records and 100 infrasound records from blides and mining explosions.

Data Split:
50% Training, 10% Validation, 40% Testing.

Baseline for Comparison:
A traditional pipeline using a STA/LTA detector for picking and an SVM classifier with hand-crafted features (duration, frequency, etc.).

Key Performance Metrics (KPIs):

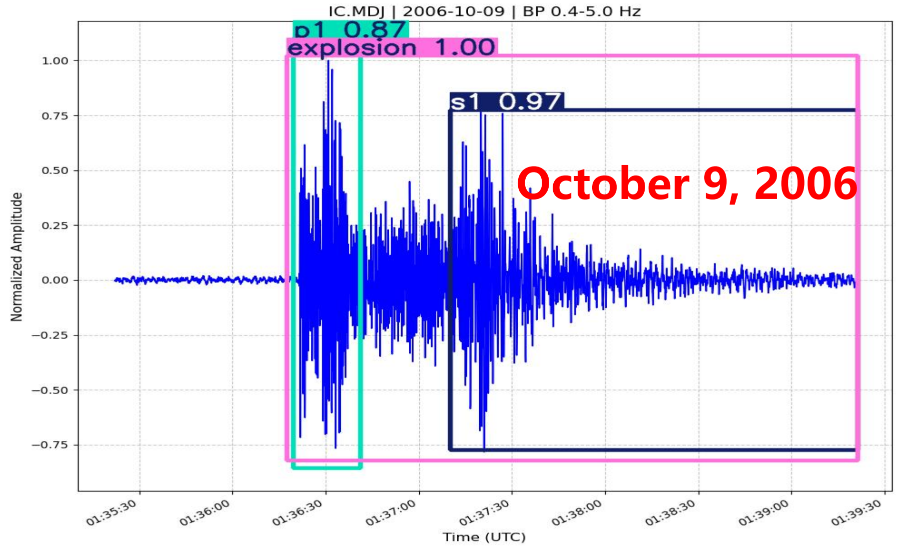
Average Precision (AP): The primary metric for object detection (area under the Precision-Recall curve).

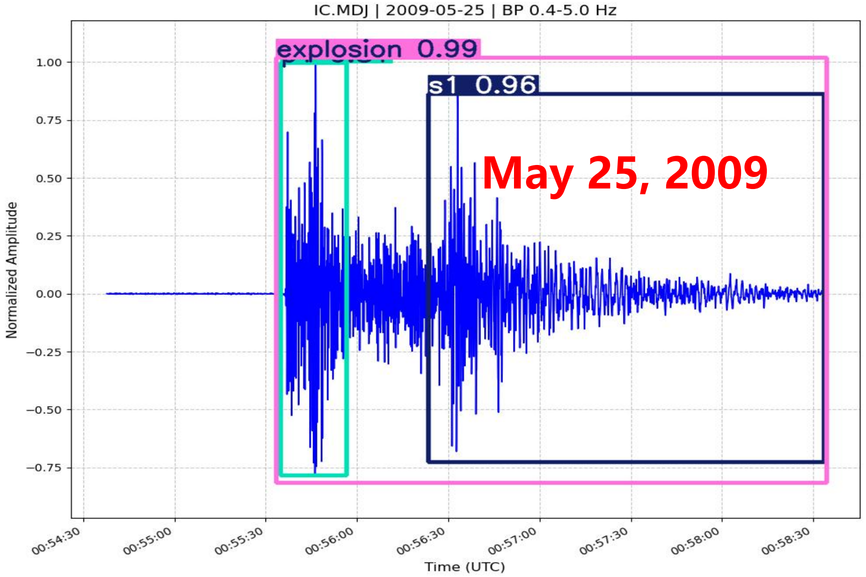
Precision: How many of the detected signals were correct?

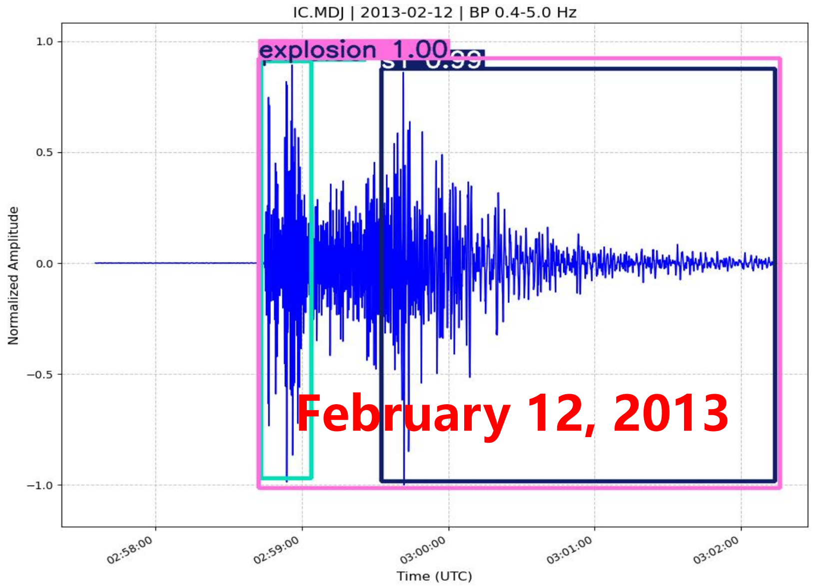
Recall: How many of the true signals did we find?

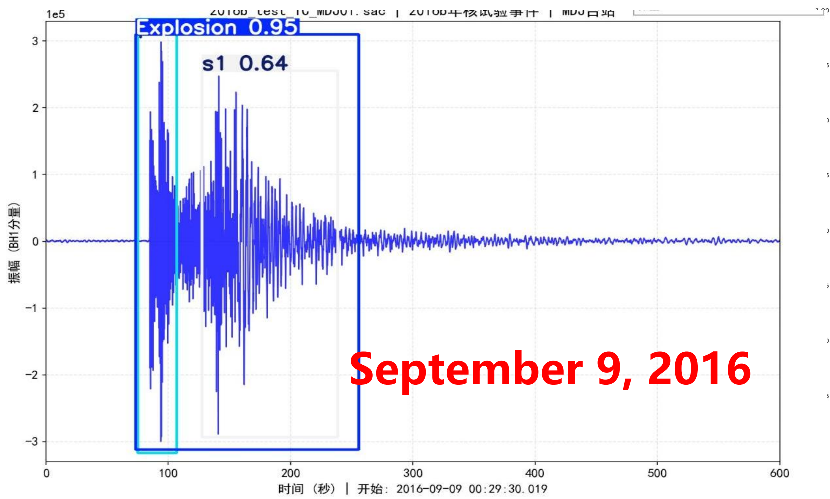
F1-Score: The harmonic mean of Precision and Recall.

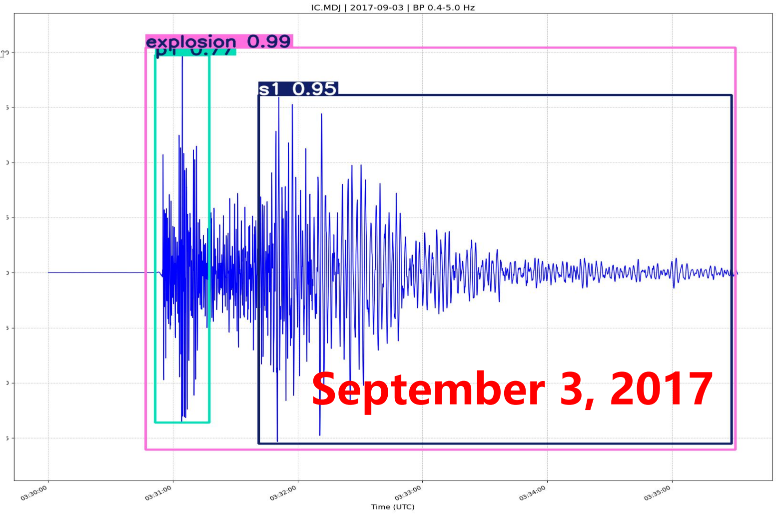
4.2 Preliminary results: Quantitative Performance (seismic signals)







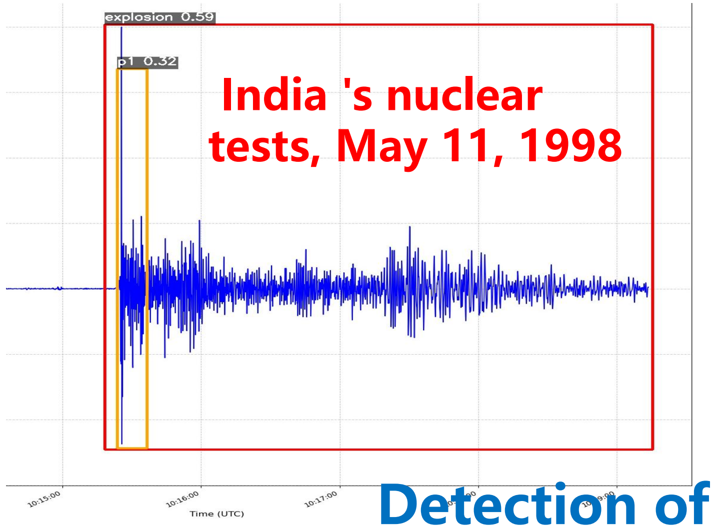


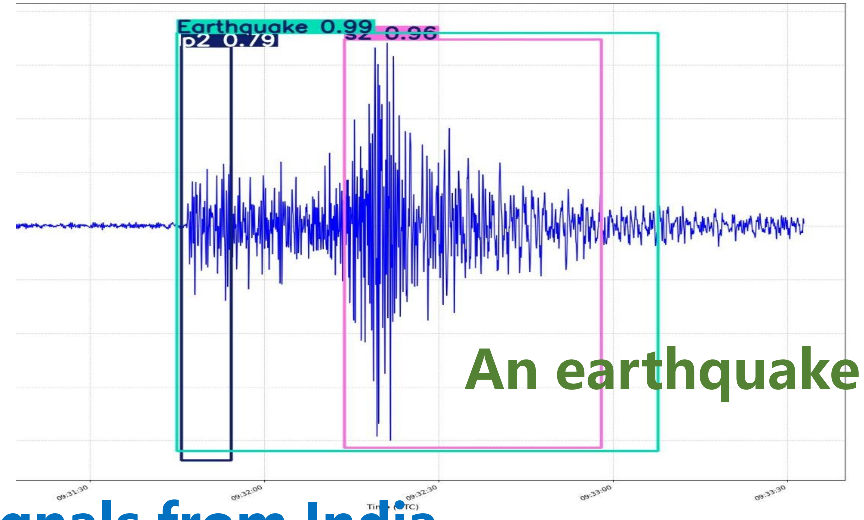


Seismic signals detection of North Korea's nuclear tests

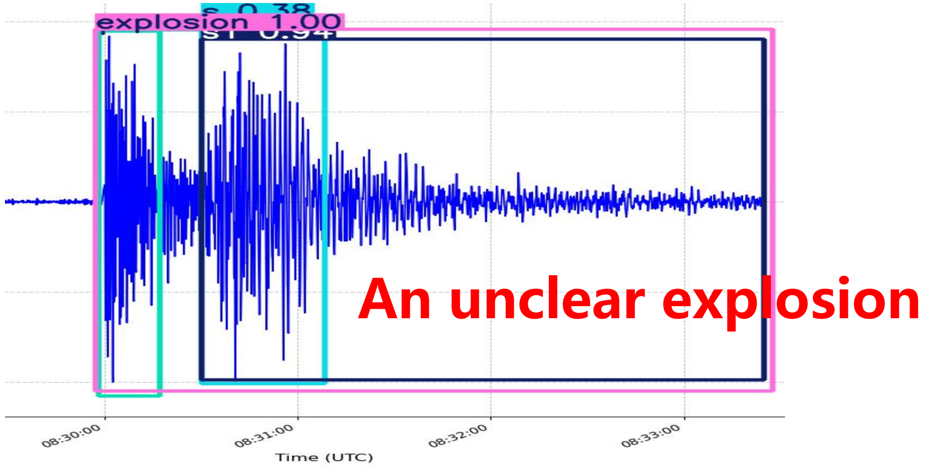
4.2 Preliminary results: Quantitative Performance

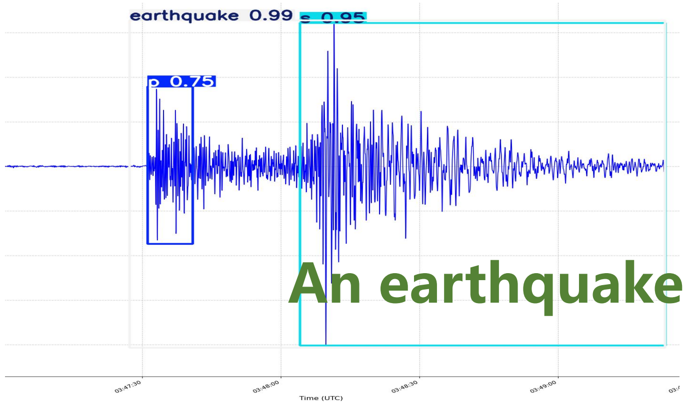
(seismic signals)





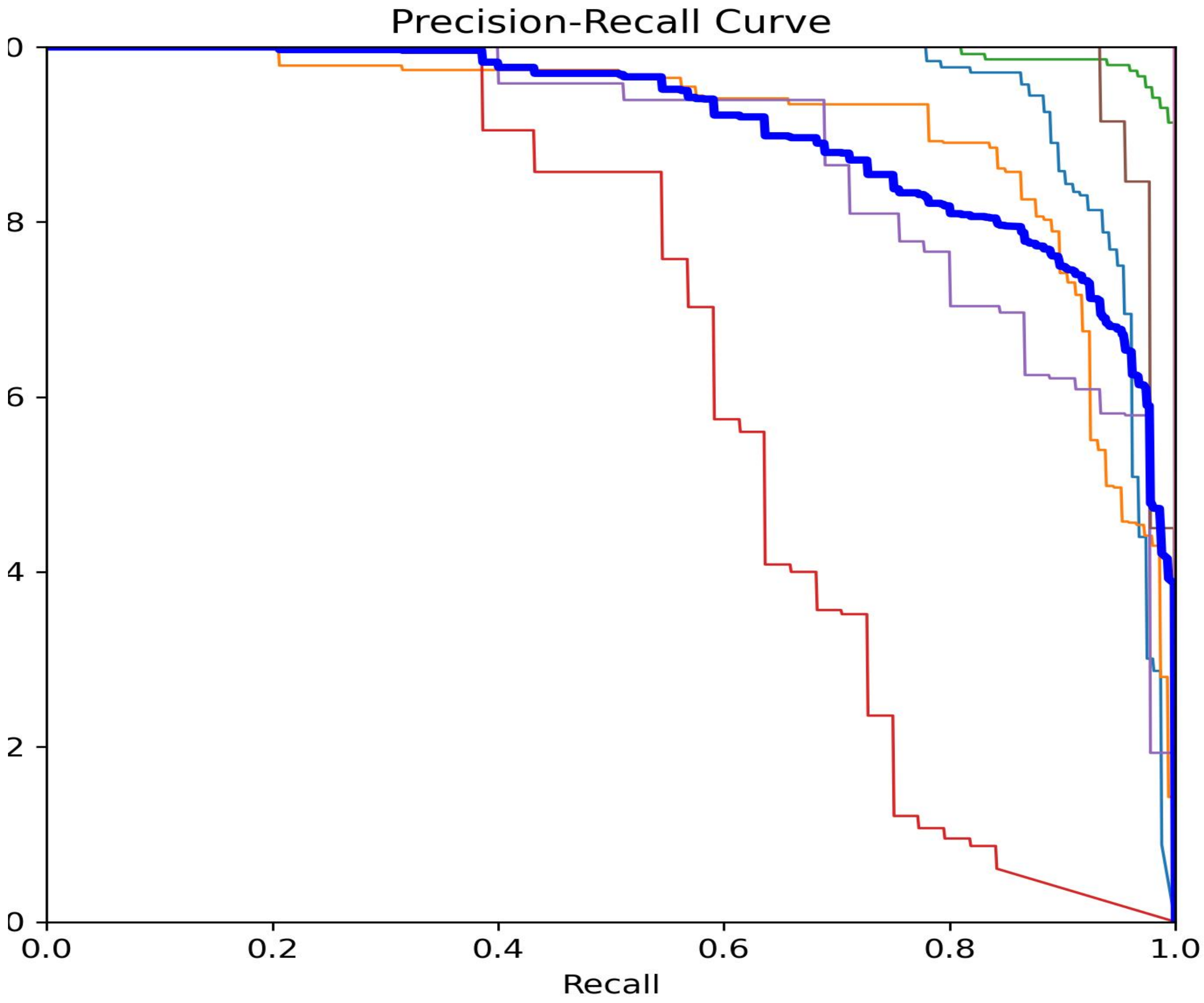
Detection of Seismic signals from India





Detection of seismic signals from Pakistan

4.2: Preliminary Results: Quantitative Performance



A clear, bold table showing the results. Use color to highlight the superior performance of the proposed method.

Evaluation Metric	Traditional Method (STA/LTA + SVM)	Proposed Method (YOLOv11)
Explosion Precision	82.2%	**97.9%**
P-wave Average Precision (AP)	81.3%	**95.5%**
S-wave Average Precision (AP)	77.1%	**91.0%**
I-wave Average Precision (AP)	74.3%	**85.4%**
Overall F1-Score	0.79	**0.85**

Significant Performance Improvement Across All Metrics

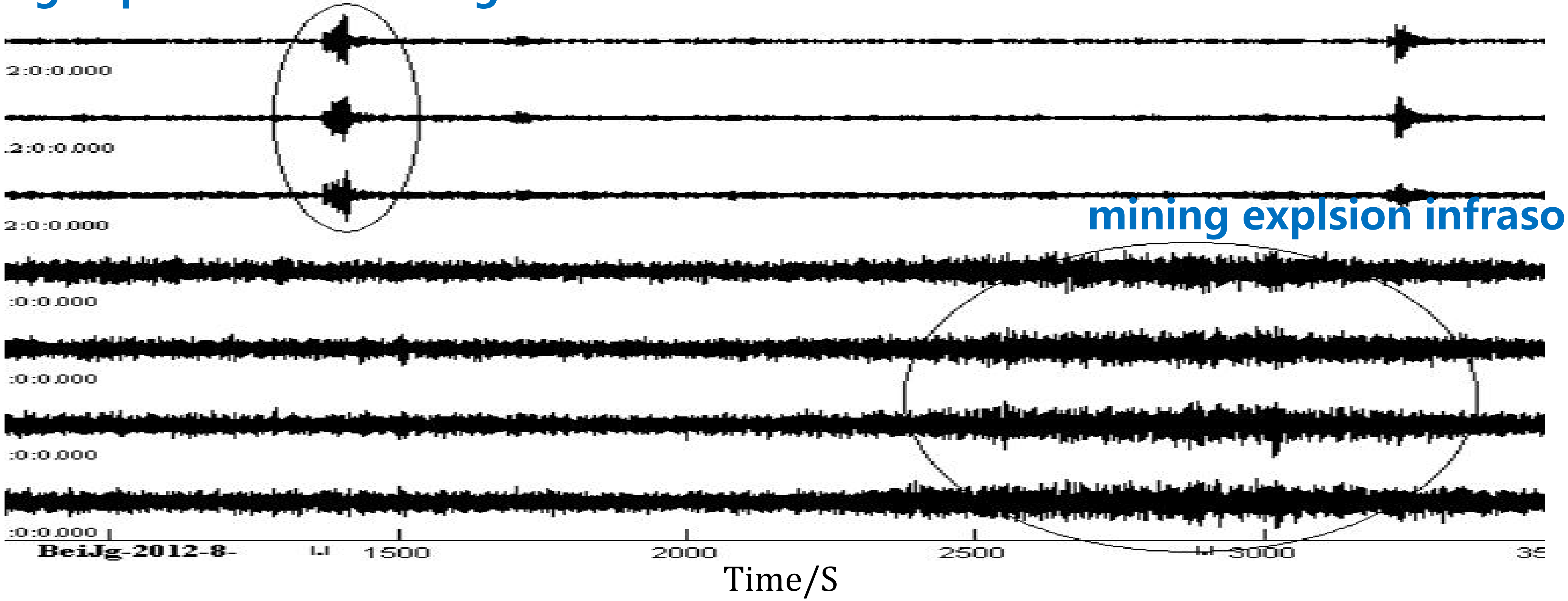
The machine vision approach demonstrates a dramatic increase in accuracy and reliability, effectively closing the performance gap with human analysts.

The results stand out:

Low-SNR performance: we tested it on a noisy seismic signal where the P-wave was barely visible to human analysts—yet the model detected it This proves it’s learned the deep, context-driven visual features that experts use (e.g., subtle frequency spikes in spectrograms).

4.2: Preliminary Results: Quantitative Performance
(infrasonic signals)

mining explosion seismic signals

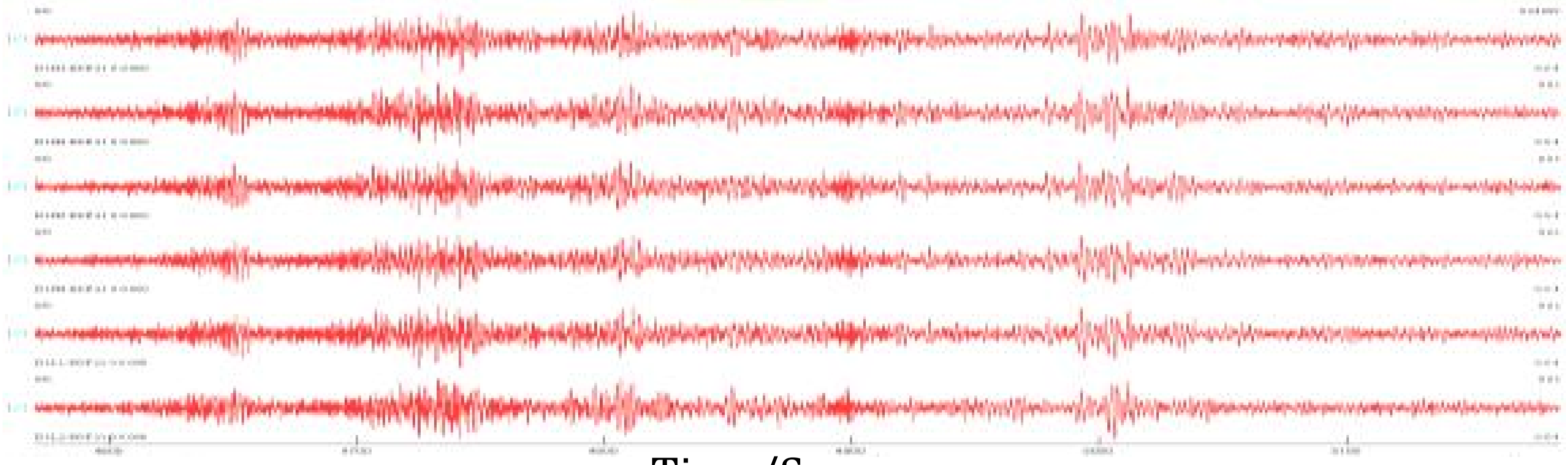


the SNR of signals is too low to identify them

4.2: Preliminary Results: Quantitative Performance
(infrasonic signals)

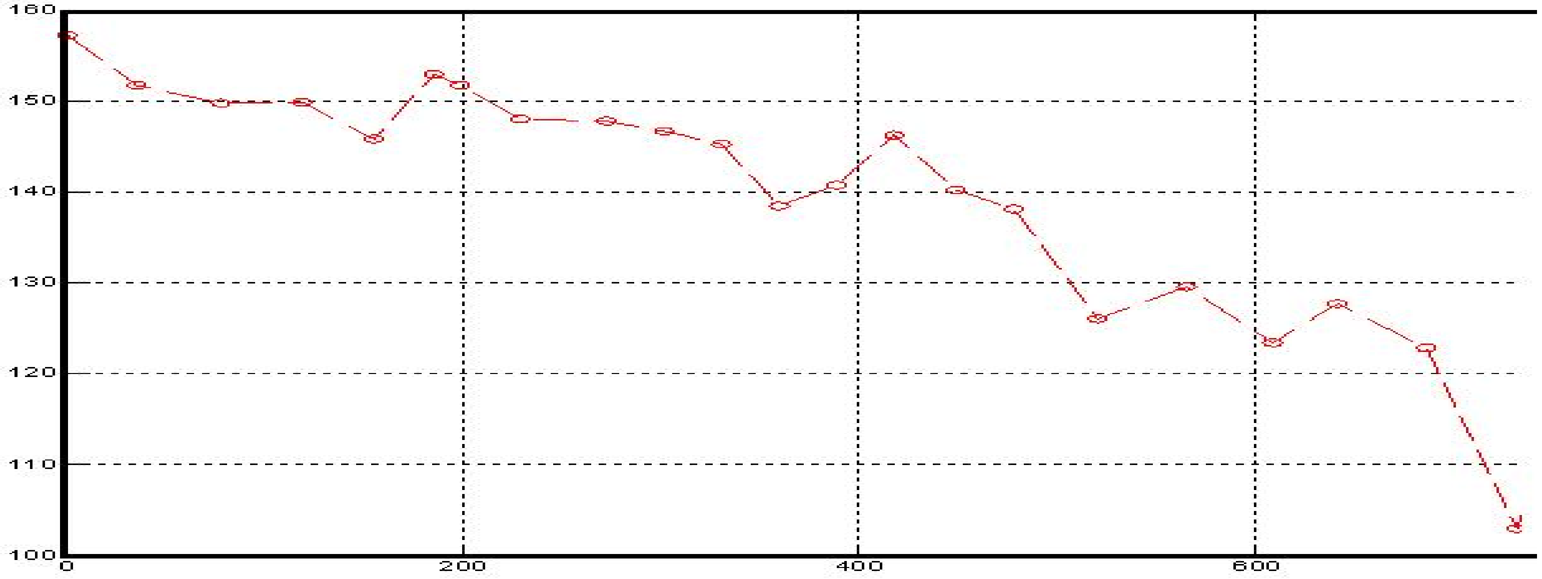
Auto-detection for Suspicious rocket launching infrasound from Baykonur and ezimuth estimation

I31KZ



Time/S

Azimuth/deg

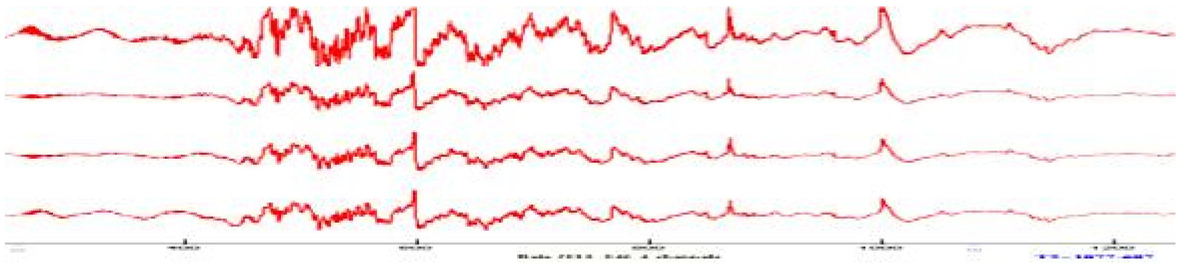


Time/S

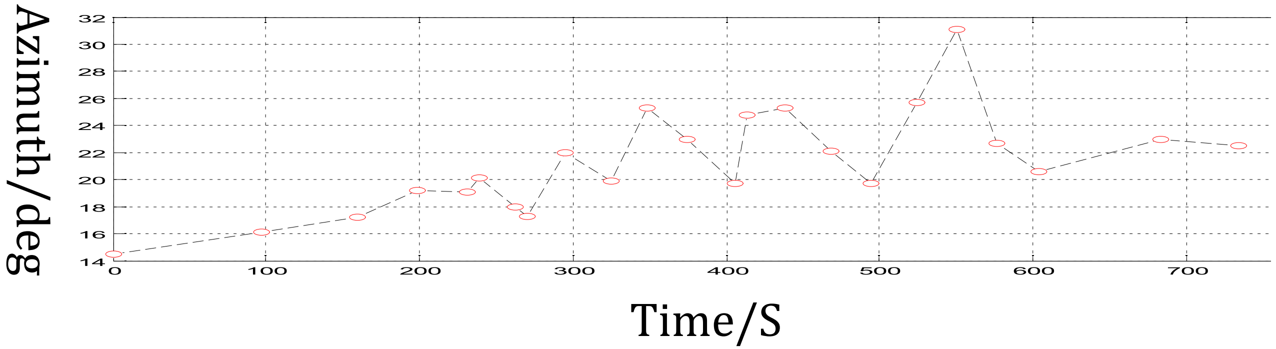
Azimuth variation vs. time

4.2: Preliminary Results: Quantitative Performance
(infrasonic signals)

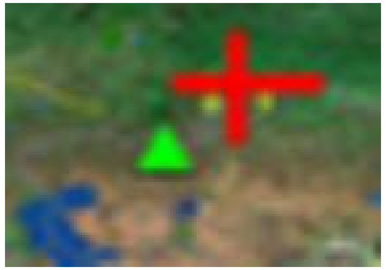
Auto-detection for the big meteor explosion in atmosphere in Russia on Feb. 15 2013



I31KZ

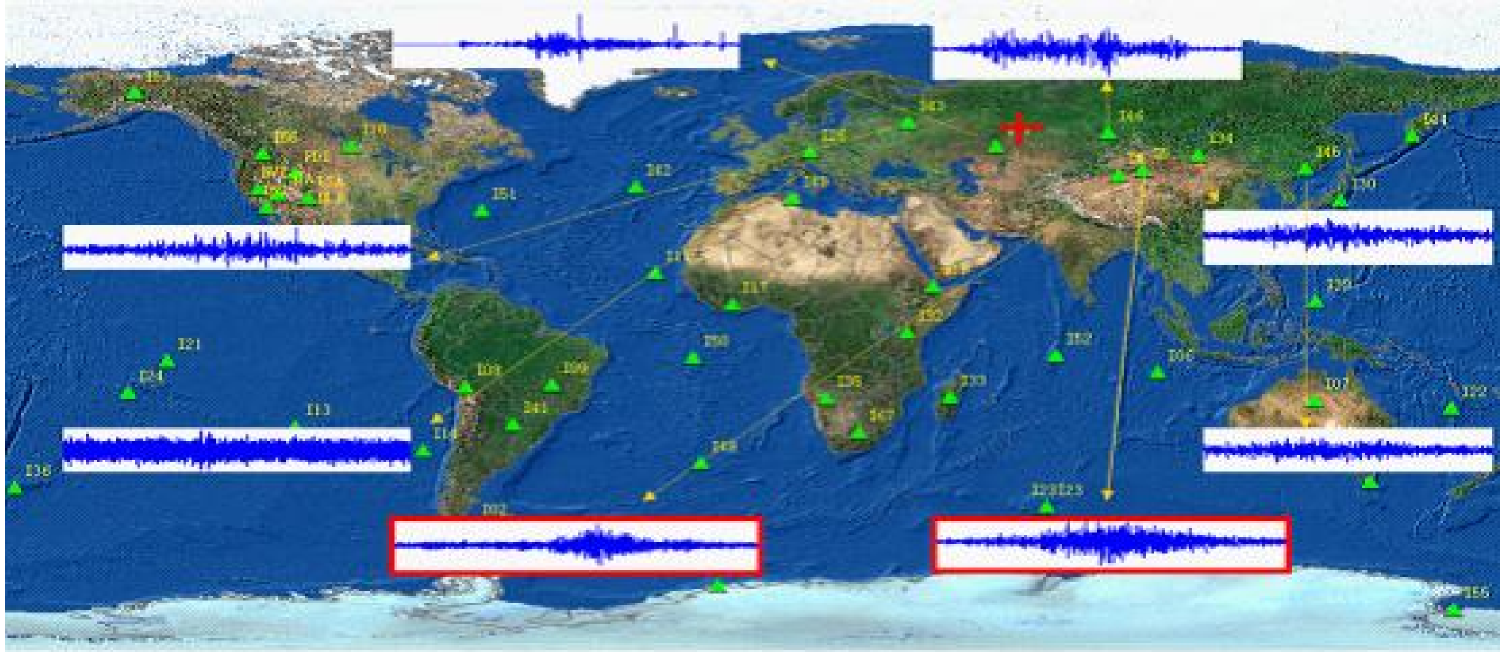


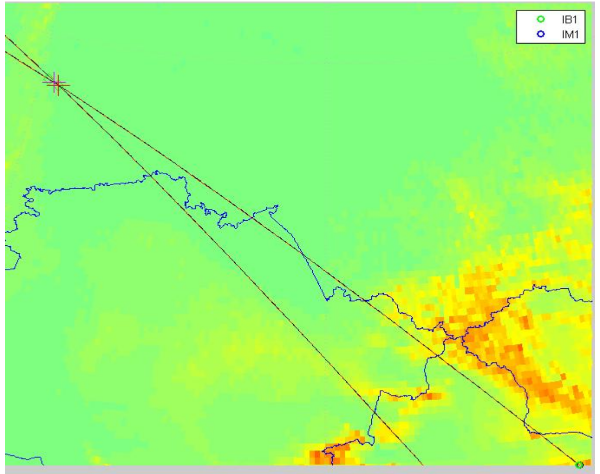
Azimuth variation vs. time

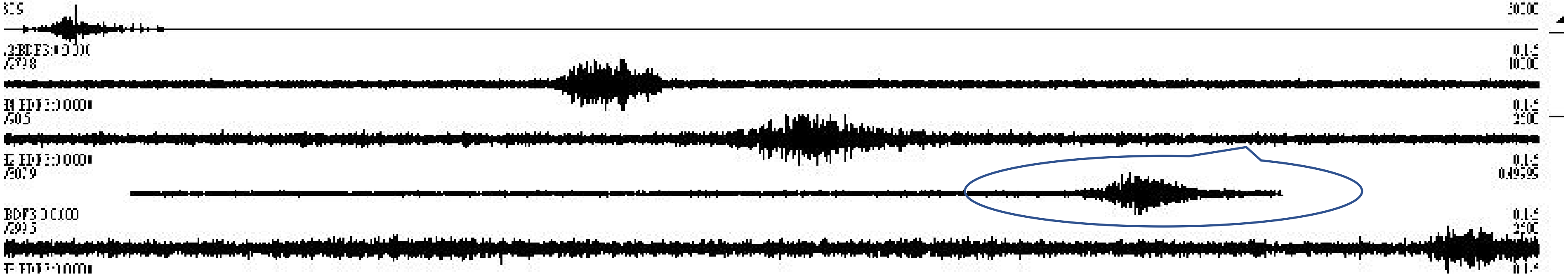


The supersonic phenomena can be observed

Detection of infrasonic signals of the big meteor explosion in atmosphere in Russia on Feb. 15，2013.







5. Conclusion & Future Work

Conclusion

What we did: We introduced a novel paradigm shift by reformulating 1D signal detection as a 2D computer vision task.

How we did it: We developed a pipeline to convert signals to images and trained a YOLOv11 model to detect and classify them.

Why it matters: This approach successfully brides the paradigm gap, allowing the transfer of human visual expertise to an automated system.

The outcome: The method significantly outperforms traditional techniques, offering superior accuracy, robustness in low-SNR conditions, and integrated detection+localization.

5. Conclusion & Future Work

Future Work

- Apply this method to large-scale data for testing and validation.
- Multi-Modal Fusion+Expert System: Train a single model on fused seismic + infrasound images to improve detection confidence and event discrimination.
- Model Lightweighting: Optimize the model for deployment on edge devices at remote monitoring stations for real-time processing.
- Explainable AI (XAI): Incorporate techniques to visualize what features the model is looking at in the image, building greater trust in its decisions.

In some aspects, the language capabilities of large models even surpass those of humans. What about image models ?

Acknowledgments

Thank you to my colleagues, and research team at Xi'an Eurasia University.

Thank you to all data providers, e.g., IRIS,CTBTO.


THANKS