The 19th International Conference on Data Mining (ICDM'19)
2019/11/8-2019/11/11, Beijing

Aftershock Detection with Multi-Scale Description based Neural Network

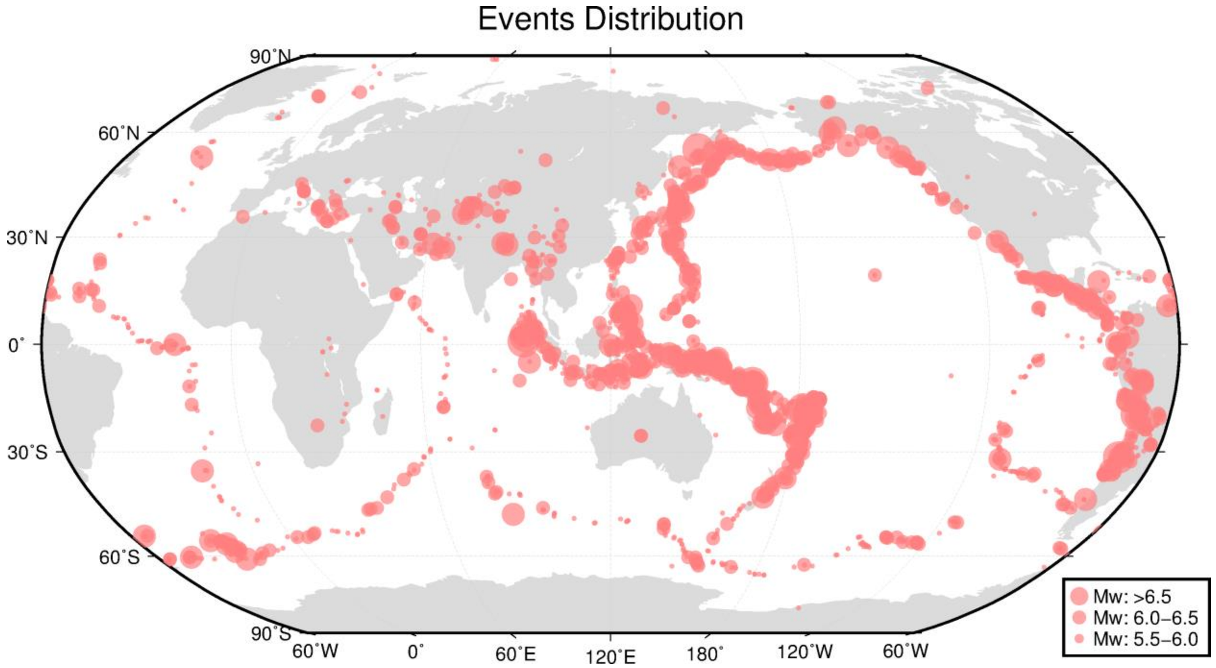
Qi Zhang[1,2], Tong Xu[1,*], Hengshu Zhu[2,*], Lifu Zhang[1], Hui Xiong[1,2,3],
Enhong Chen[1], Qi Liu[1]

[1]Anhui Province Key Lab of Big Data Analysis and Application,
University of Science and Technology of China
[2]Baidu Talent Intelligence Center, Baidu Inc,
[3]Business Intelligence Lab, Baidu Research

Background

◆ Earthquake Events

Earthquake is one kind of worst nature disasters which may cause injury and loss of life and collapse of buildings.



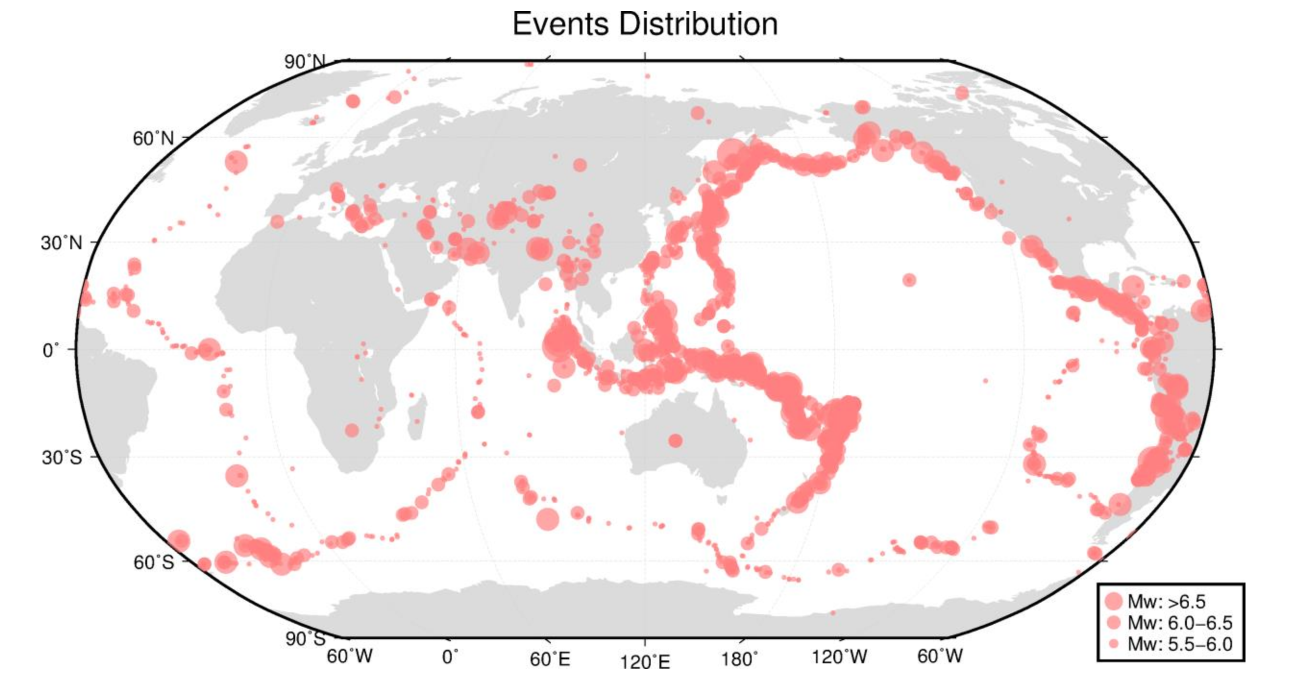
Earthquake events distribution all over the world in 2018

Nature Disaster: Earthquake

Background

◆ Earthquake Events

Earthquake is one kind of worst nature disasters which may cause injury and loss of life and collapse of buildings.



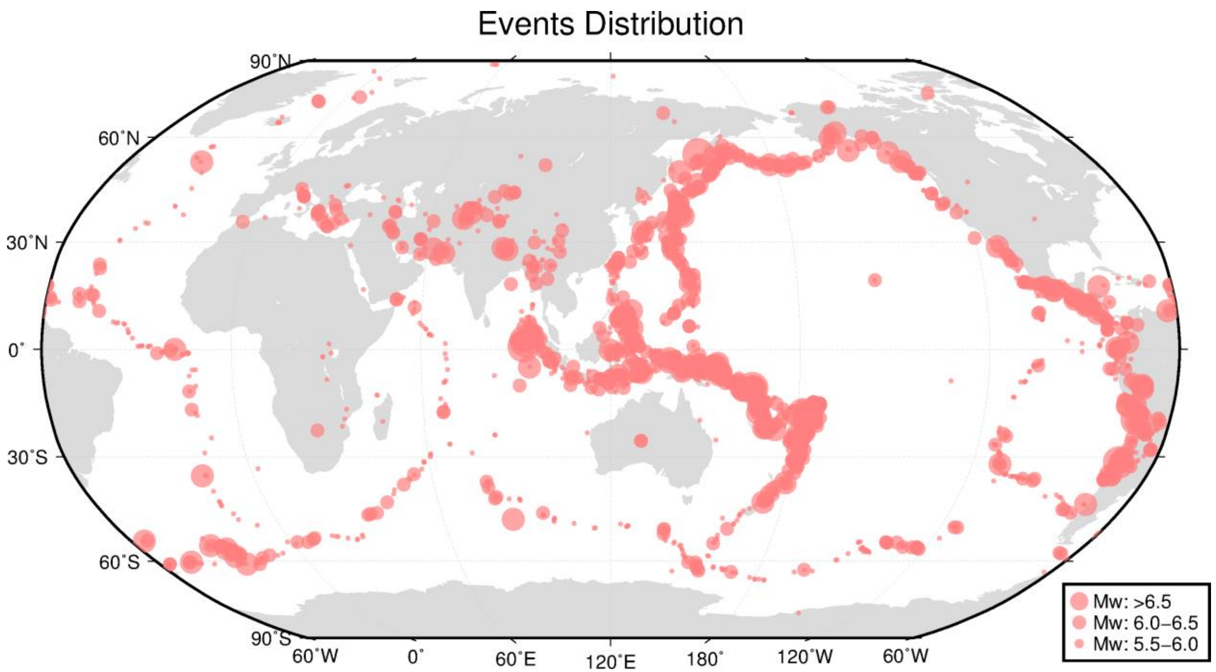
Earthquake events distribution all over the world in 2018

Nature Disaster: Earthquake

Background

◆ Earthquake Events

Earthquake is one kind of worst nature disasters which may cause injury and loss of life and collapse of buildings.



Earthquake events distribution all over the world in 2018

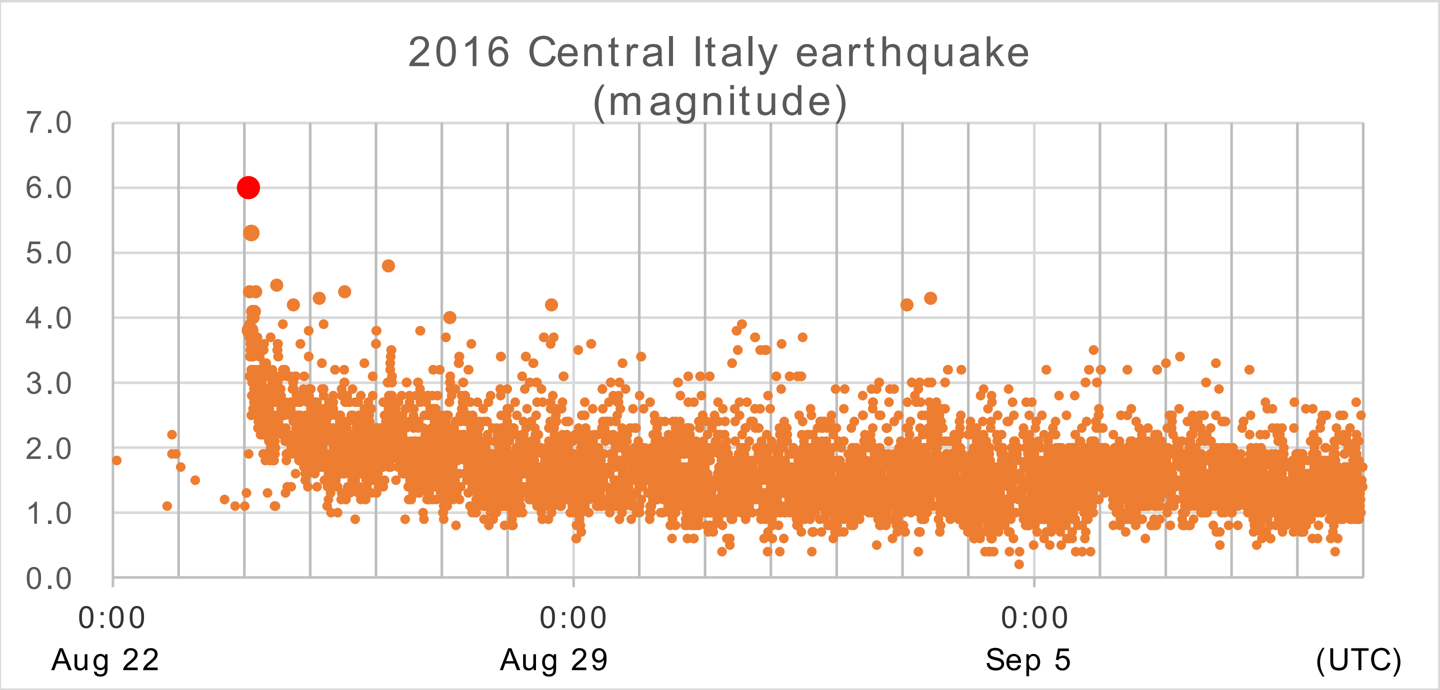
Nature Disaster: Earthquake

Background

Aftershock

Aftershocks refer to the smaller earthquakes that occur following large earthquakes, in the same area of the main shock.

- Difficult to detect.
- Collapse buildings that are damaged from the main shock.

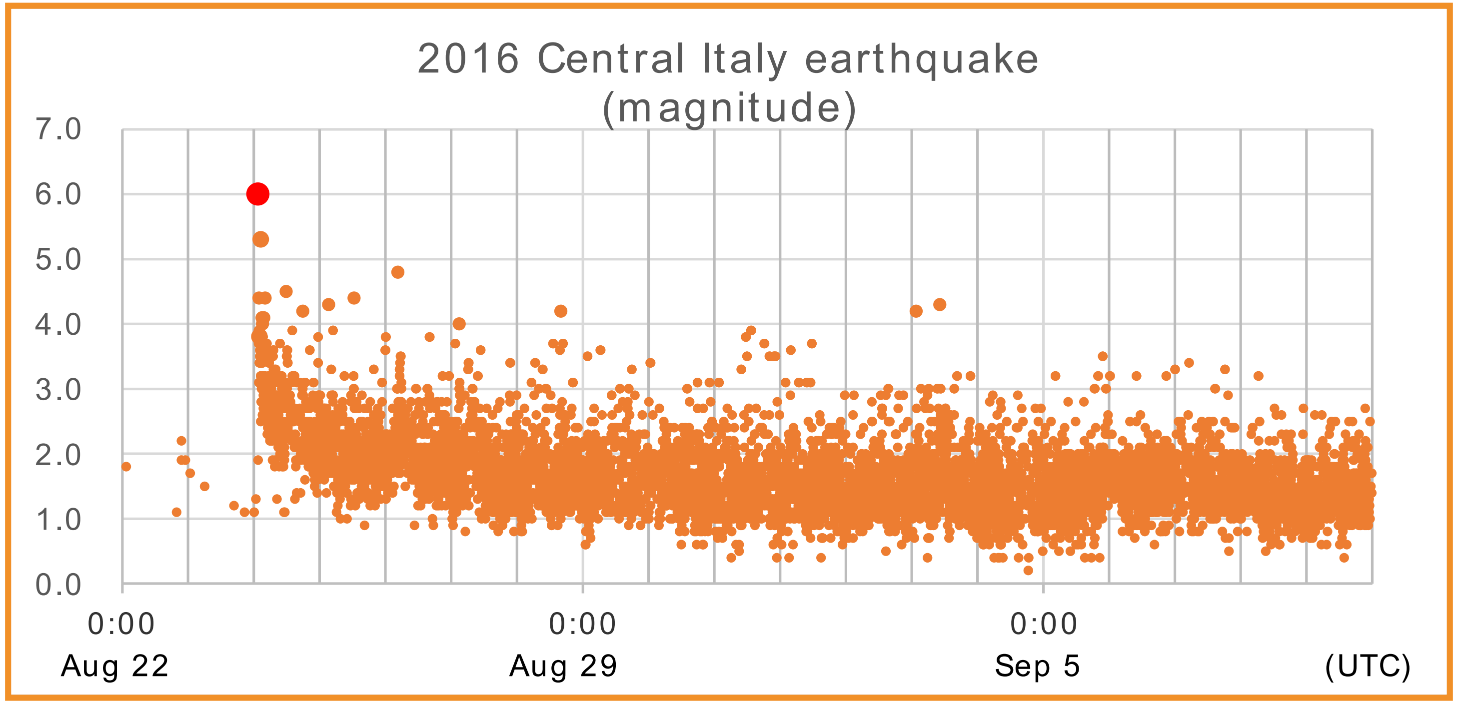


Background

Aftershock

Aftershocks refer to the smaller earthquakes that occur following large earthquakes, in the same area of the main shock.

- Difficult to detect.
- Collapse buildings that are damaged from the main shock.

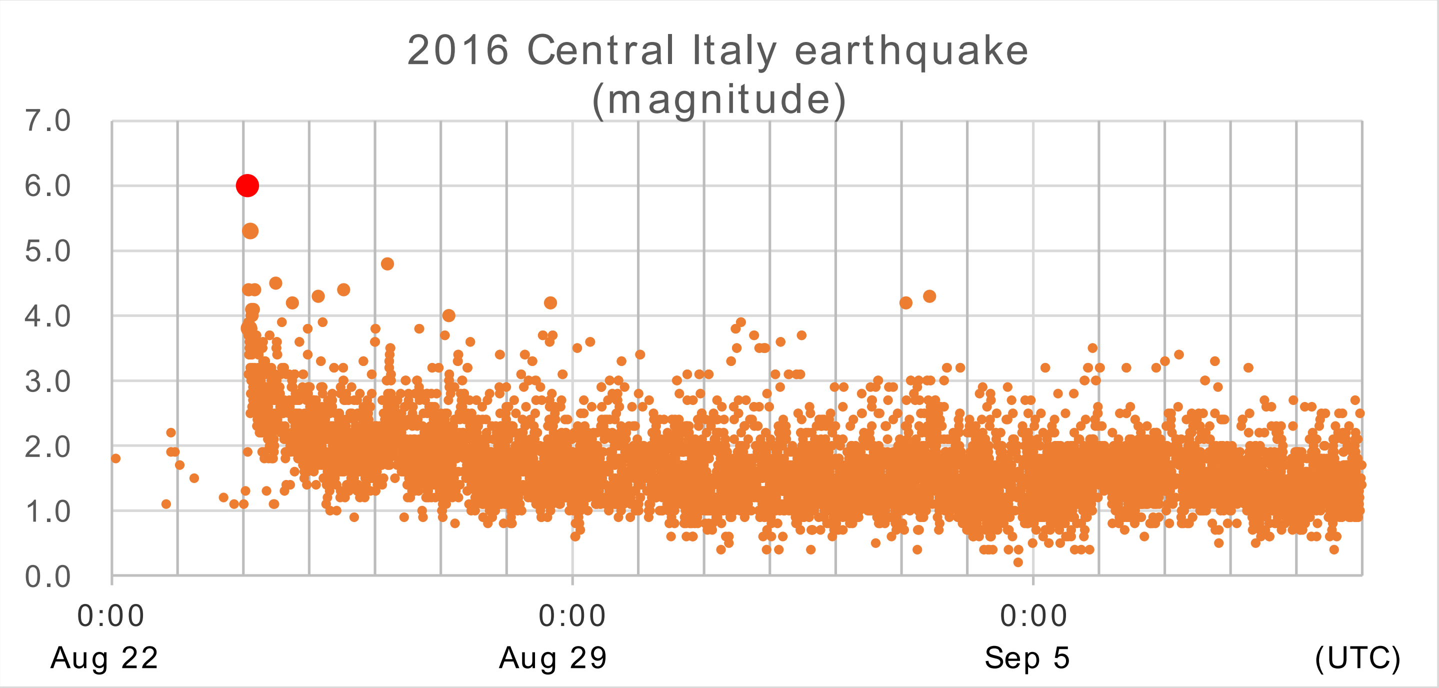


Background

- **Aftershock**

Aftershocks refer to the smaller earthquakes that occur following large earthquakes, in the same area of the main shock.

- Difficult to detect.
- Collapse buildings that are damaged from the main shock.

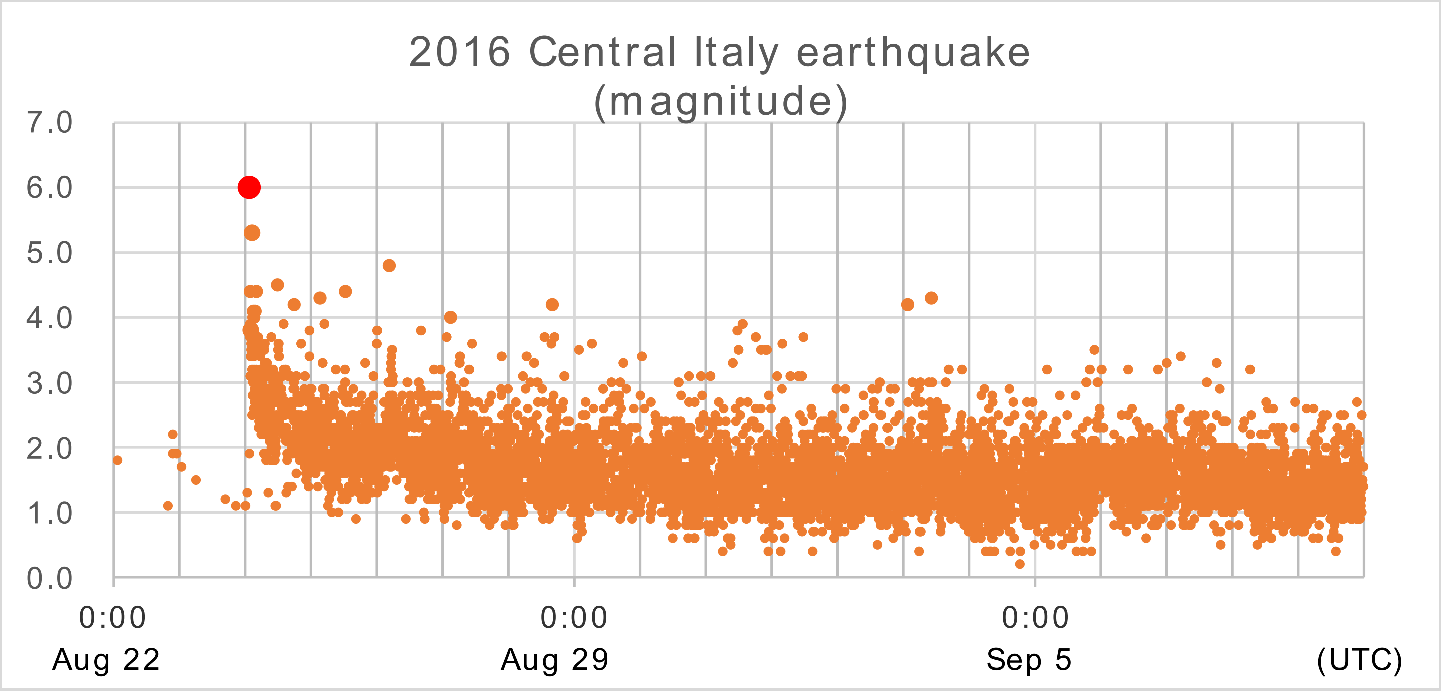


Background

◆ Aftershock

Aftershocks refer to the smaller earthquakes that occur following large earthquakes, in the same area of the main shock.

- Difficult to detect.
- Collapse buildings that are damaged from the main shock.



Background

- **Motivation**

Benefit

- Automatic aftershock detection can support emergence actions.

- Useful for the research of geological activity and seismic mechanism.

Challenges

- Highly-noise and weak signal. How to effectively detect aftershocks?
  Multi-Scale Description based Neural Network.

- Multiple monitoring stations. How to utilize this relationship?
  multi-task learning strategy .

Background

Motivation

Benefit

- Automatic aftershock detection can support emergence actions.

- Useful for the research of geological activity and seismic mechanism.

Challenges

- Highly-noise and weak signal. How to effectively detect aftershocks?
  Multi-Scale Description based Neural Network.

- Multiple monitoring stations. How to utilize this relationship?
  multi-task learning strategy .

Background

◆ Motivation

Benefit

- Automatic aftershock detection can support emergence actions.
- Useful for the research of geological activity and seismic mechanism.

Challenges

- Highly-noise and weak signal. How to effectively detect aftershocks?
  Multi-Scale Description based Neural Network.
- Multiple monitoring stations. How to utilize this relationship?
  multi-task learning strategy .

Motivation

Benefit

- Automatic aftershock detection can support emergence actions.
- Useful for the research of geological activity and seismic mechanism.

Challenges

- Highly-noise and weak signal. How to effectively detect aftershocks?
  Multi-Scale Description based Neural Network.
- Multiple monitoring stations. How to utilize this relationship?
  multi-task learning strategy .

Problem Formulation

◆ Aftershock Detection

Definition: Machine-learning-based Aftershock Detection.

Given a set of waveform windows D, where each $d_i \in D$ has a label l_i for indicating the existence of seismic P-wave, the **objective** is to learn a predictive model M for classifying waveform windows with respect to the label y_i.

Problem Formulation

◆ Aftershock Detection

Definition: Machine-learning-based Aftershock Detection.

Given a set of waveform windows D, where each $d_i \in D$ has a label l_i for indicating the existence of seismic P-wave, the **objective** is to learn a predictive model M for classifying waveform windows with respect to the label y_i.

Problem Formulation

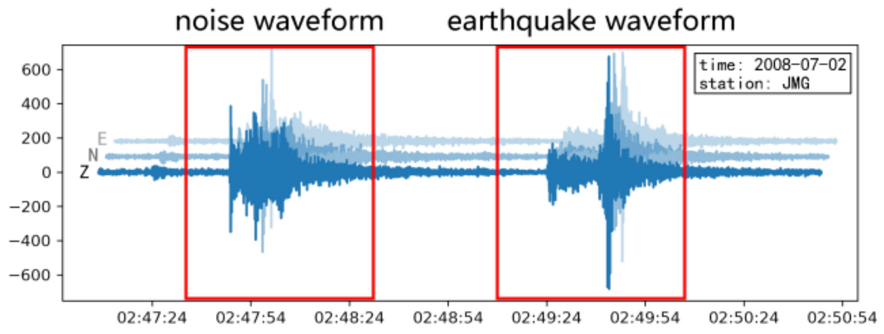
Aftershock Detection

Definition: Machine-learning-based Aftershock Detection.

Given a set of waveform windows D, where each $d_i \in D$ has a label l_i for indicating the existence of seismic P-wave, the **objective** is to learn a predictive model M for classifying waveform windows with respect to the label y_i.

Methodology

- **Data Set Description**
  - The waveforms after the Wenchuan M8.0 Earthquake.
  - There are 2,833 aftershocks, corresponding to 9,891 pieces of seismic waveforms in short time window.
  - These waveforms were recorded in three spatial dimensions (i.e., Z for the vertical channel, N for the north-south channel, and E for the east-west channel) by 15 monitoring stations.



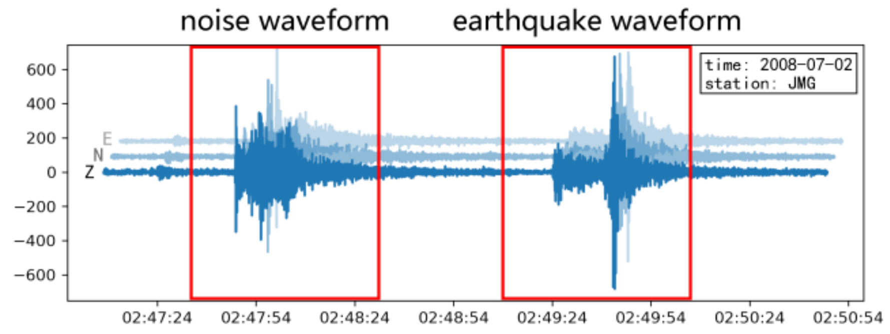
(a) Example for aftershock detection.

TABLE I: The number of aftershock waveforms.

Station	Number	Homology Number	Percent
JMG	1,208	1,208	100%
YZP	1,072	1,015	94.7%
QCH	894	890	99.6%
PUW	1,350	1,338	99.1%
WXT	839	838	99.9%
SPA	574	574	100%
XJI	614	612	99.7%
HSH	821	821	100%
YGD	166	166	100%
JJS	908	903	99.4%
MXI	1,215	1,196	98.4%
XCO	223	223	100%
WDT	6	6	100%
MIAX	1	1	100%
SUM	9,891	9,791	99.0%

Methodology

- ## Data Set Description
  - The waveforms after the Wenchuan M8.0 Earthquake.
  - There are 2,833 aftershocks, corresponding to 9,891 pieces of seismic waveforms in short time window.
  - These waveforms were recorded in three spatial dimensions (i.e., Z for the vertical channel, N for the north-south channel, and E for the east-west channel) by 15 monitoring stations.



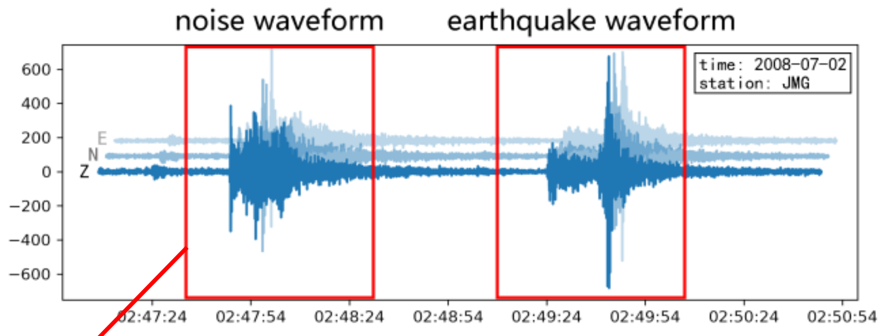
(a) Example for aftershock detection.

TABLE I: The number of aftershock waveforms.

Station	Number	Homology Number	Percent
JMG	1,208	1,208	100%
YZP	1,072	1,015	94.7%
QCH	894	890	99.6%
PUW	1,350	1,338	99.1%
WXT	839	838	99.9%
SPA	574	574	100%
XJI	614	612	99.7%
HSH	821	821	100%
YGD	166	166	100%
JJS	908	903	99.4%
MXI	1,215	1,196	98.4%
XCO	223	223	100%
WDT	6	6	100%
MIAX	1	1	100%
SUM	9,891	9,791	99.0%

Methodology

- **Data Set Description**
  - The waveforms after the Wenchuan M8.0 Earthquake.
  - There are 2,833 aftershocks, corresponding to 9,891 pieces of seismic waveforms in short time window.
  - These waveforms were recorded in three spatial dimensions (i.e., Z for the vertical channel, N for the north-south channel, and E for the east-west channel) by 15 monitoring stations.



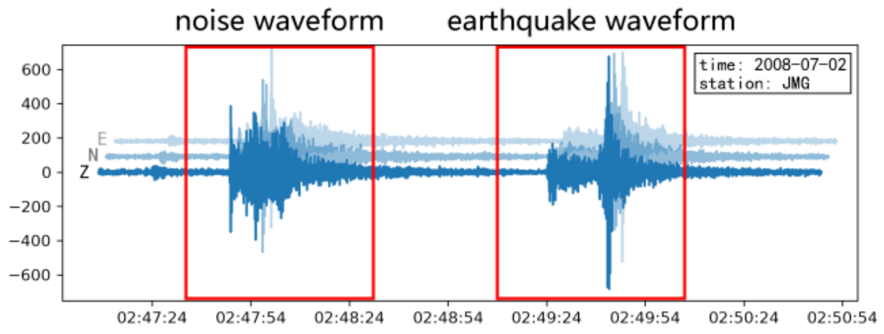
(a) Example for aftershock detection.

TABLE I: The number of aftershock waveforms.

Station	Number	Homology Number	Percent
JMG	1,208	1,208	100%
YZP	1,072	1,015	94.7%
QCH	894	890	99.6%
PUW	1,350	1,338	99.1%
WXT	839	838	99.9%
SPA	574	574	100%
XJI	614	612	99.7%
HSH	821	821	100%
YGD	166	166	100%
JJS	908	903	99.4%
MXI	1,215	1,196	98.4%
XCO	223	223	100%
WDT	6	6	100%
MIAX	1	1	100%
SUM	9,891	9,791	99.0%

Methodology

- **Data Set Description**
  - The waveforms after the Wenchuan M8.0 Earthquake.
  - There are 2,833 aftershocks, corresponding to 9,891 pieces of seismic waveforms in short time window.
  - These waveforms were recorded in three spatial dimensions (i.e., Z for the vertical channel, N for the north-south channel, and E for the east-west channel) by 15 monitoring stations.



(a) Example for aftershock detection.

TABLE I: The number of aftershock waveforms.

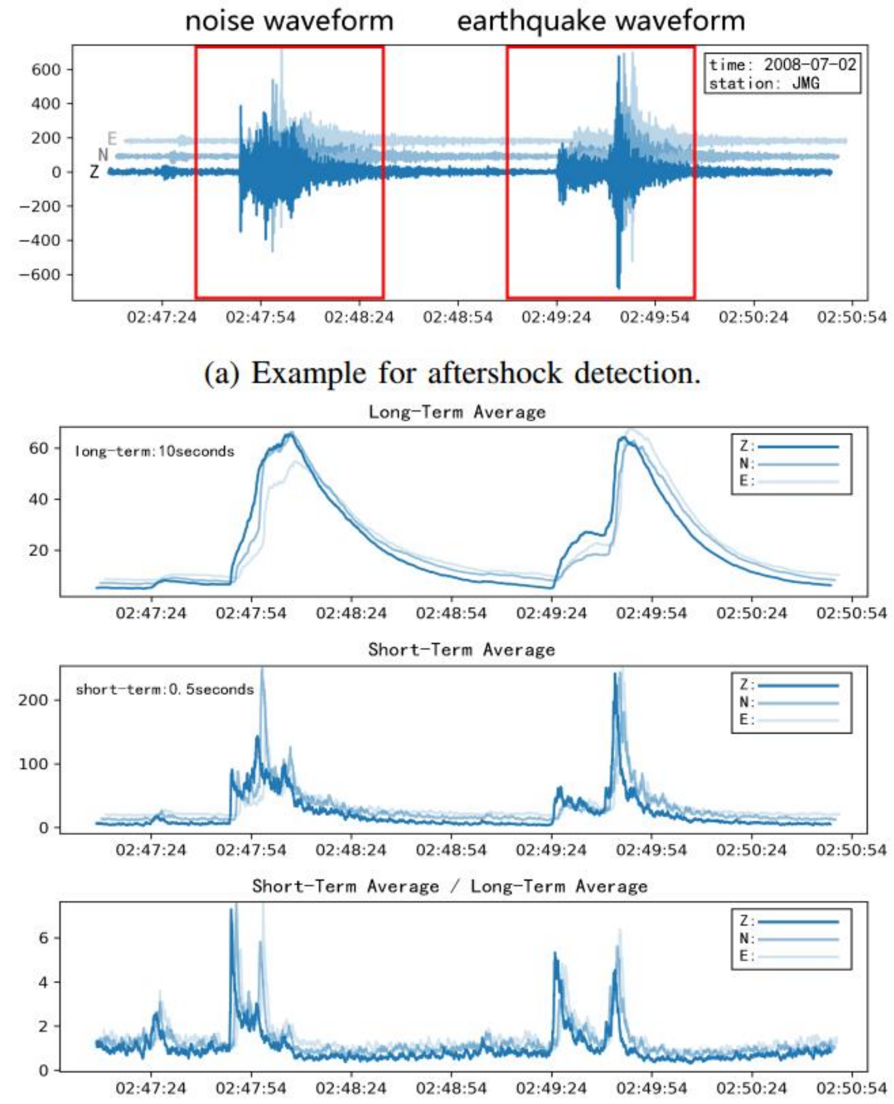
Station	Number	Homology Number	Percent
JMG	1,208	1,208	100%
YZP	1,072	1,015	94.7%
QCH	894	890	99.6%
PUW	1,350	1,338	99.1%
WXT	839	838	99.9%
SPA	574	574	100%
XJI	614	612	99.7%
HSH	821	821	100%
YGD	166	166	100%
JJS	908	903	99.4%
MXI	1,215	1,196	98.4%
XCO	223	223	100%
WDT	6	6	100%
MIAX	1	1	100%
SUM	9,891	9,791	99.0%

Methodology

- **Characteristics of the Data**

  - STA/LTA(Short-Term Average/Long-Term Average) is the most widely-used earthquake detection approach.

  - The different scale-aware descriptions reflect different characteristics of seismic waveform.



(a) Example for aftershock detection.

(b) Example of classical multi-scale description.

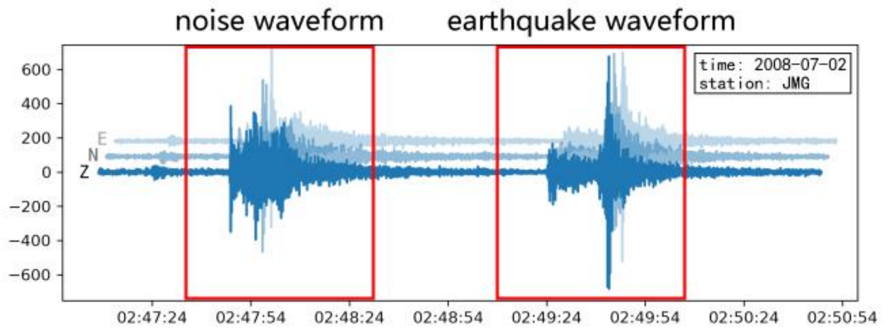
Fig. 2: Some motivating examples of our multi-scale description based aftershock detection approach.

Methodology

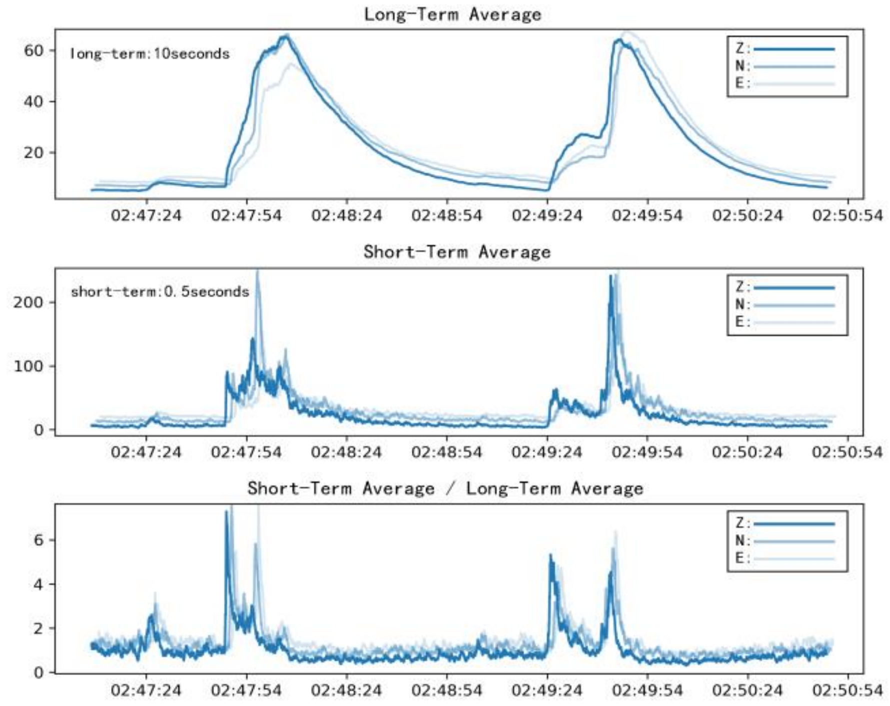
- ## Characteristics of the Data

  - STA/LTA(Short-Term Average/Long-Term Average) is the most widely-used earthquake detection approach.

  - The different scale-aware descriptions reflect different characteristics of seismic waveform.



(a) Example for aftershock detection.



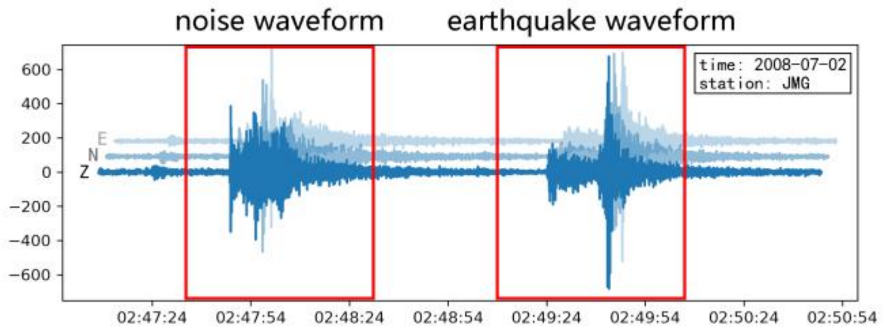
(b) Example of classical multi-scale description.

Fig. 2: Some motivating examples of our multi-scale description based aftershock detection approach.

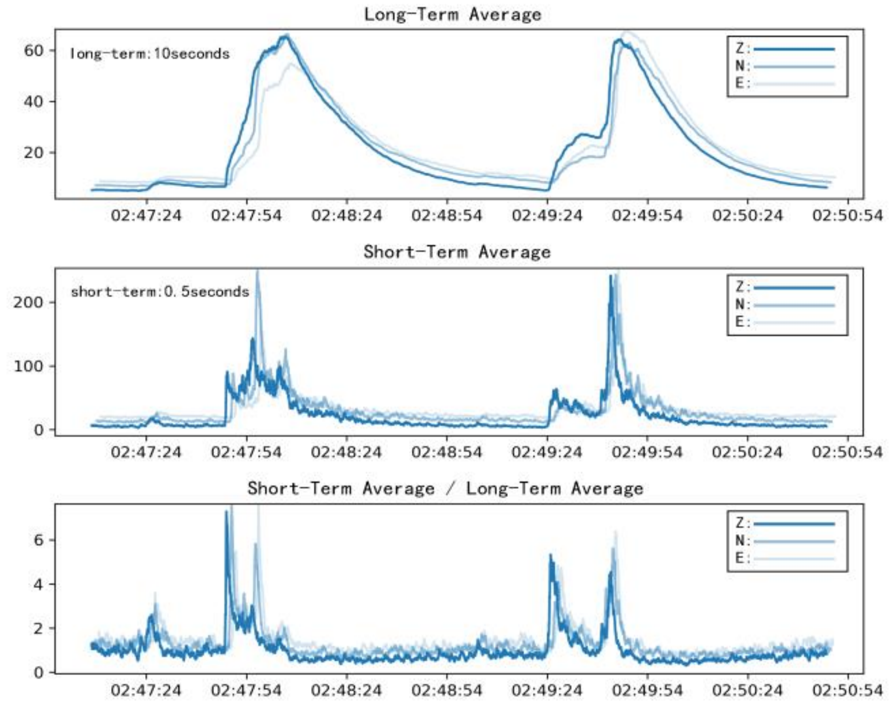
Methodology

- ## Characteristics of the Data

We design a **Multi-Scale Description Based Neural Network** to extract appropriate features for improving the performance of aftershock detection.



(a) Example for aftershock detection.



(b) Example of classical multi-scale description.

Fig. 2: Some motivating examples of our multi-scale description based aftershock detection approach.

Methodology

◆ MSD-cell: Generating Multi-Scale Description

the module needs to implement two key functions

First function: extract multi-scale description

- The first function can remember prior features on different scales and add new scale feature.
- Implemented by a 1x3x32/1 convolutions

Second function: mix multi-scale description

- the second function can compare and mix these two kinds of features.
- Implemented by a 1x1x32/1 convolutions

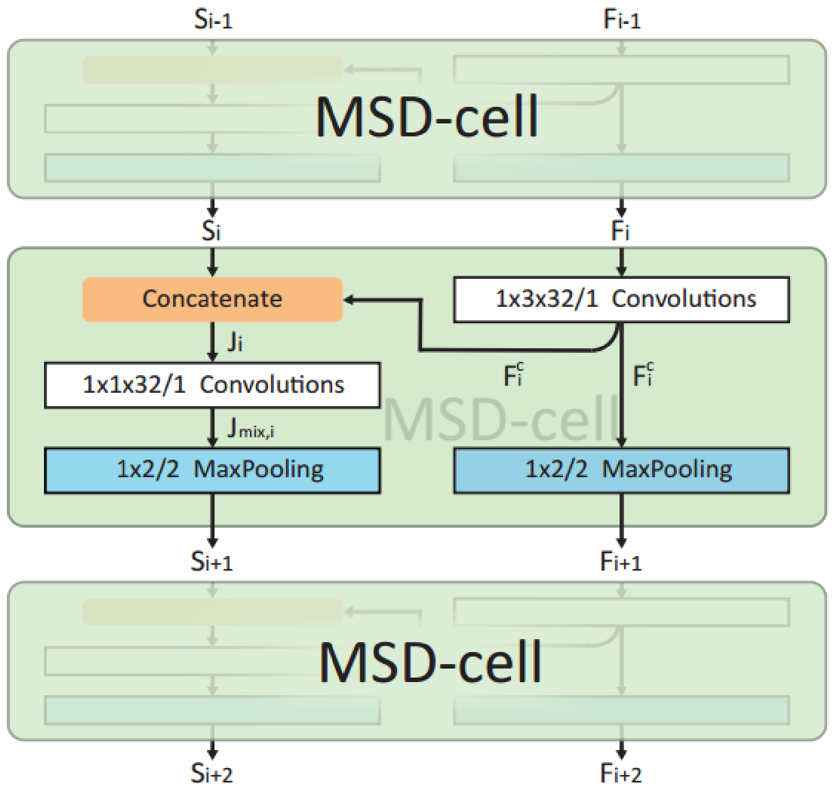


Fig. 3: The detailed structure of MSD-cell, which can be expanded easily.

MSD-cell: Generating Multi-Scale Description

the module needs to implement two key functions

First function: extract multi-scale description

- The first function can remember prior features on different scales and add new scale feature.
- Implemented by a $\boxed{\text{1x3x32/1 convolutions}}$

Second function: mix multi-scale description

- the second function can compare and mix these two kinds of features.
- Implemented by a $\boxed{\text{1x1x32/1 convolutions}}$

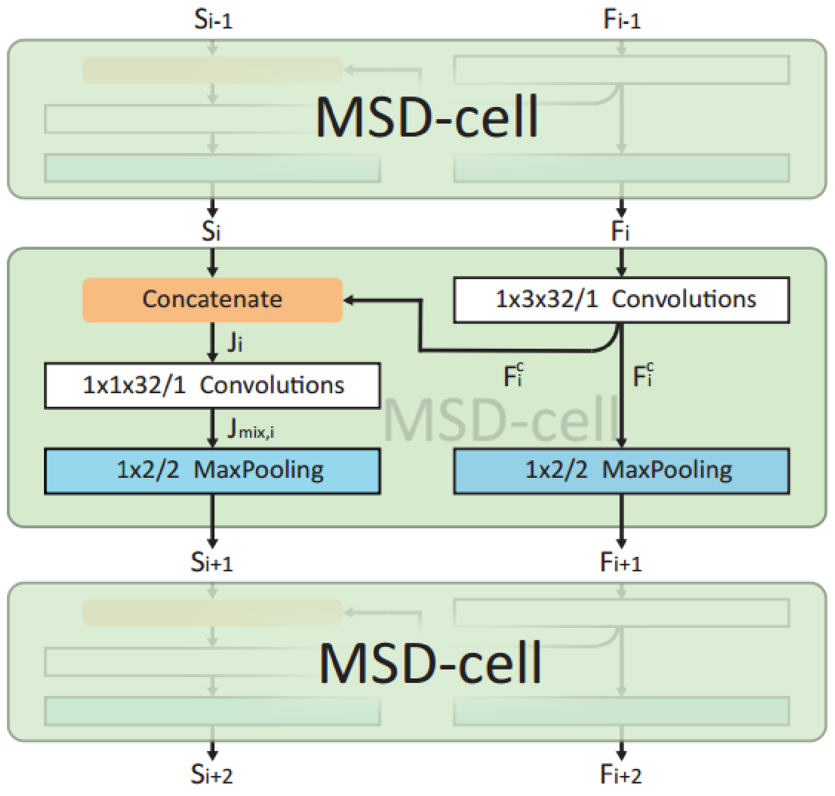


Fig. 3: The detailed structure of MSD-cell, which can be expanded easily.

Methodology

◆ MSD-cell: Generating Multi-Scale Description

the module needs to implement two key functions

First function: extract multi-scale description

- The first function can remember prior features on different scales and add new scale feature.
- Implemented by a 1x3x32/1 convolutions

Second function: mix multi-scale description

- the second function can compare and mix these two kinds of features.
- Implemented by a 1x1x32/1 convolutions

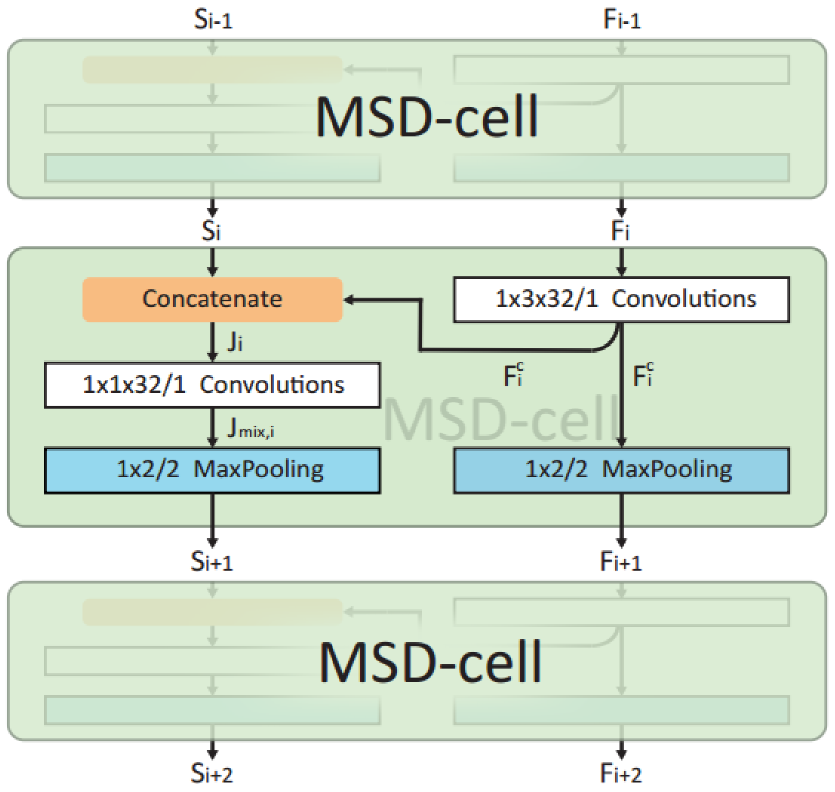


Fig. 3: The detailed structure of MSD-cell, which can be expanded easily.

Methodology

◆ MSD-cell: Generating Multi-Scale Description

the module needs to implement two key functions

First function: extract multi-scale description

- The first function can remember prior features on different scales and add new scale feature.
- Implemented by a $\boxed{\text{1x3x32/1 convolutions}}$

Second function: mix multi-scale description

- the second function can compare and mix these two kinds of features.
- Implemented by a $\boxed{\text{1x1x32/1 convolutions}}$

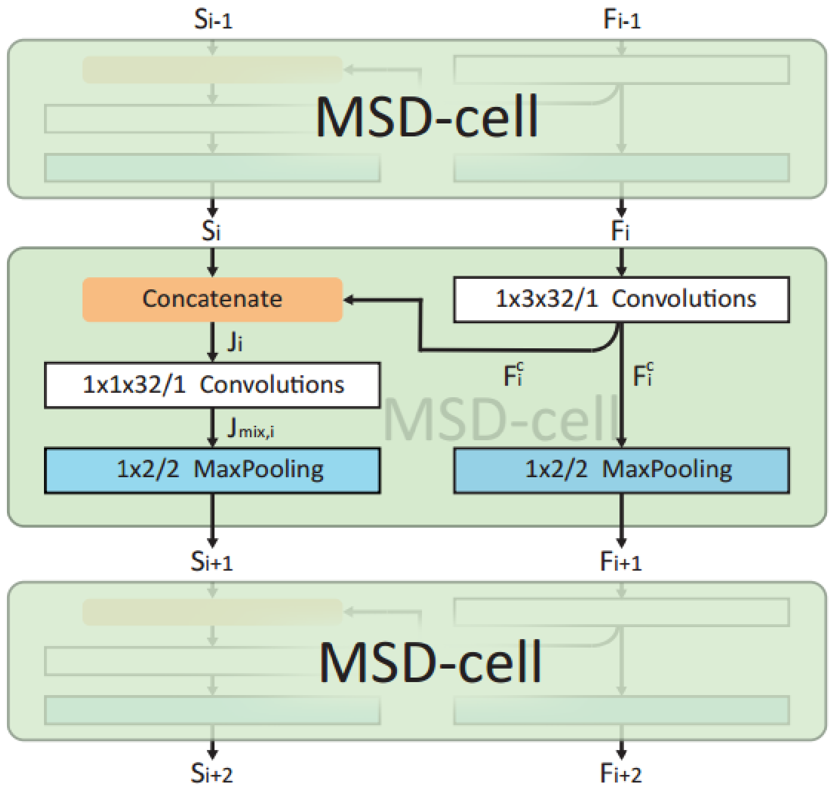


Fig. 3: The detailed structure of MSD-cell, which can be expanded easily.

Methodology

MSD-cell: Generating Multi-Scale Description

the module needs to implement two key functions

First function: extract multi-scale description

- The first function can remember prior features on different scales and add new scale feature.
- Implemented by a 1x3x32/1 convolutions

Second function: mix multi-scale description

- the second function can compare and mix these two kinds of features.
- Implemented by a 1x1x32/1 convolutions

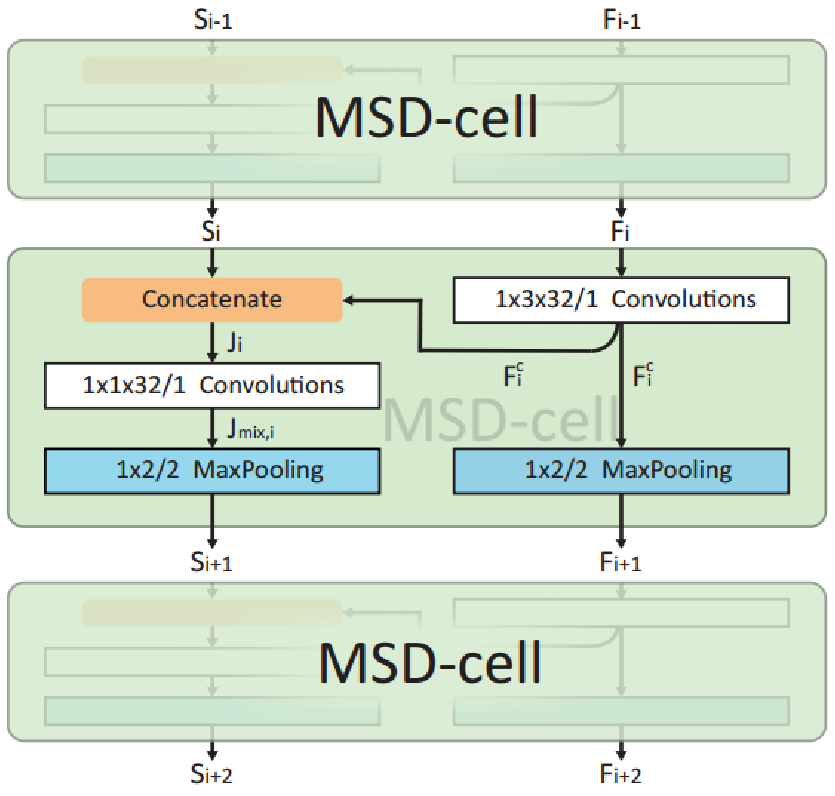


Fig. 3: The detailed structure of MSD-cell, which can be expanded easily.

Methodology

MSD-cell: Generating Multi-Scale Description

the module needs to implement two key functions

First function: extract multi-scale description

- The first function can remember prior features on different scales and add new scale feature.
- Implemented by a 1x3x32/1 convolutions

Second function: mix multi-scale description

- the second function can compare and mix these two kinds of features.
- Implemented by a 1x1x32/1 convolutions

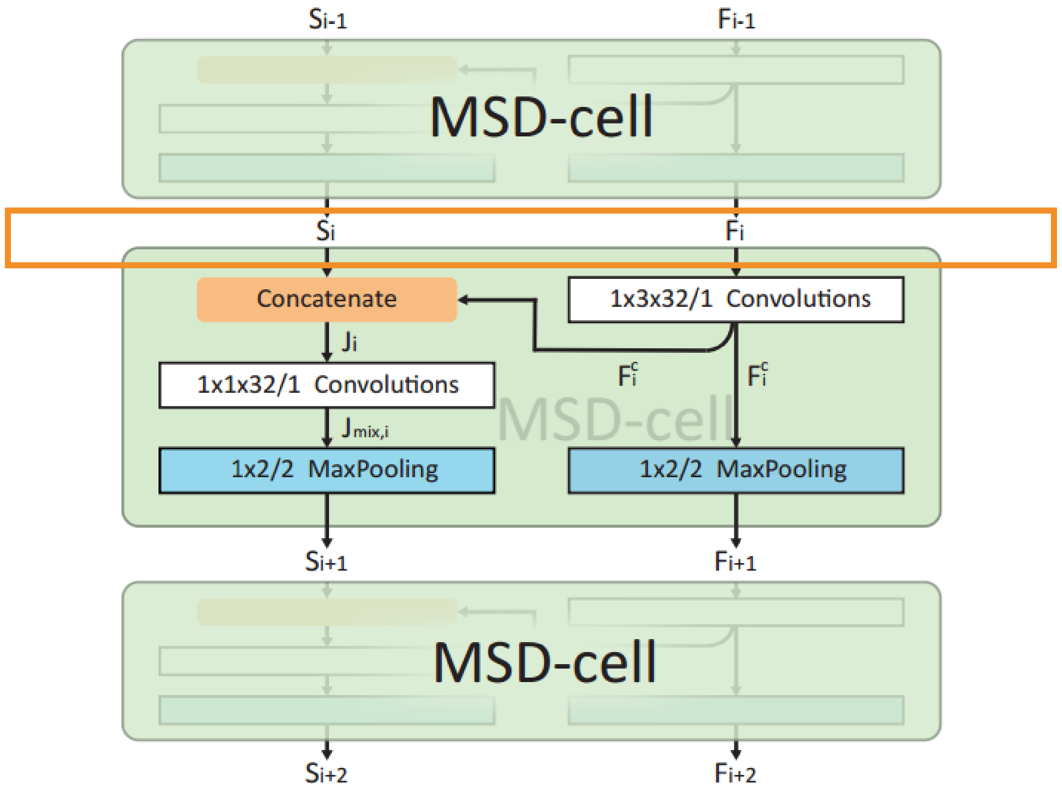


Fig. 3: The detailed structure of MSD-cell, which can be expanded easily.

MSD-cell: Generating Multi-Scale Description

the module needs to implement two key functions

First function: extract multi-scale description

- The first function can remember prior features on different scales and add new scale feature.
- Implemented by a 1x3x32/1 convolutions

Second function: mix multi-scale description

- the second function can compare and mix these two kinds of features.
- Implemented by a 1x1x32/1 convolutions

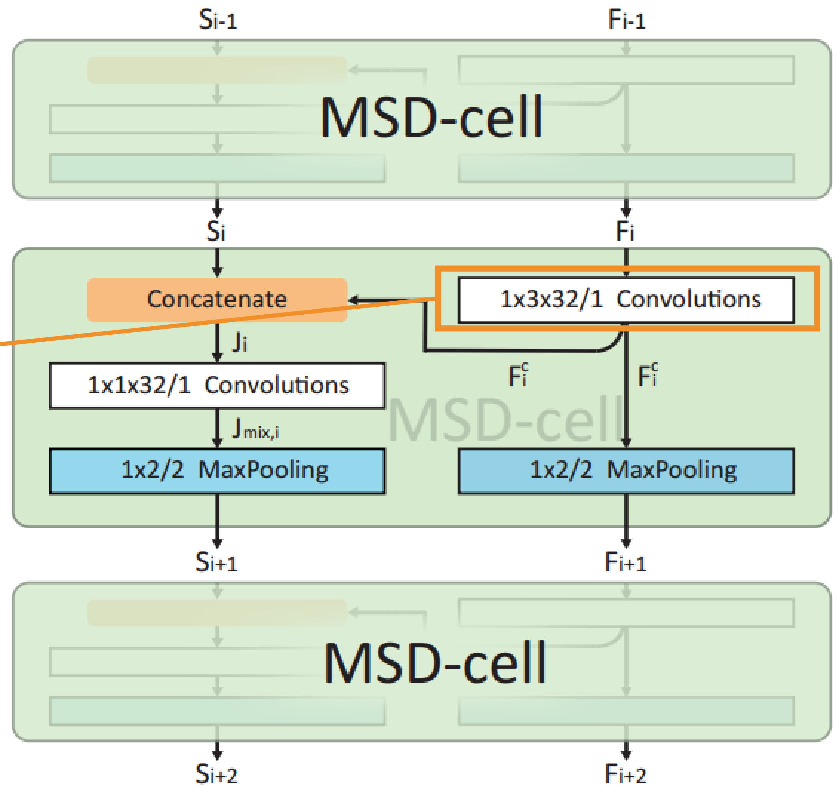


Fig. 3: The detailed structure of MSD-cell, which can be expanded easily.

MSD-cell: Generating Multi-Scale Description

the module needs to implement two key functions

First function: extract multi-scale description

- The first function can remember prior features on different scales and add new scale feature.
- Implemented by a 1x3x32/1 convolutions

Second function: mix multi-scale description

- the second function can compare and mix these two kinds of features.
- Implemented by a 1x1x32/1 convolutions

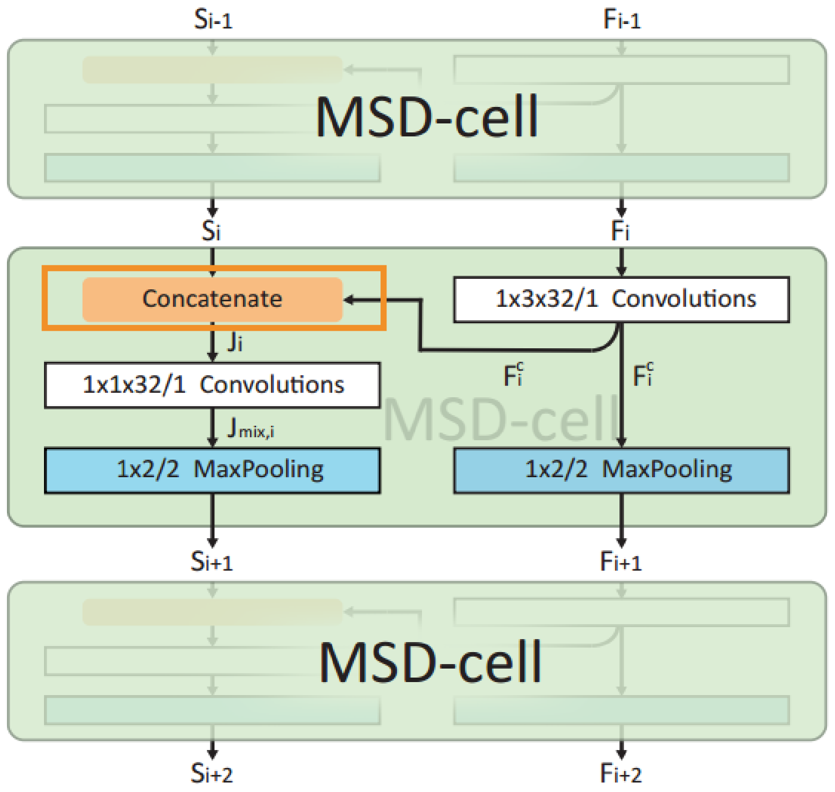


Fig. 3: The detailed structure of MSD-cell, which can be expanded easily.

Methodology

◆ **MSD-cell: Generating Multi-Scale Description**

the module needs to implement two key functions

First function: extract multi-scale description

- The first function can remember prior features on different scales and add new scale feature.
- Implemented by a 1x3x32/1 convolutions

Second function: mix multi-scale description

- the second function can compare and mix these two kinds of features.
- Implemented by a 1x1x32/1 convolutions

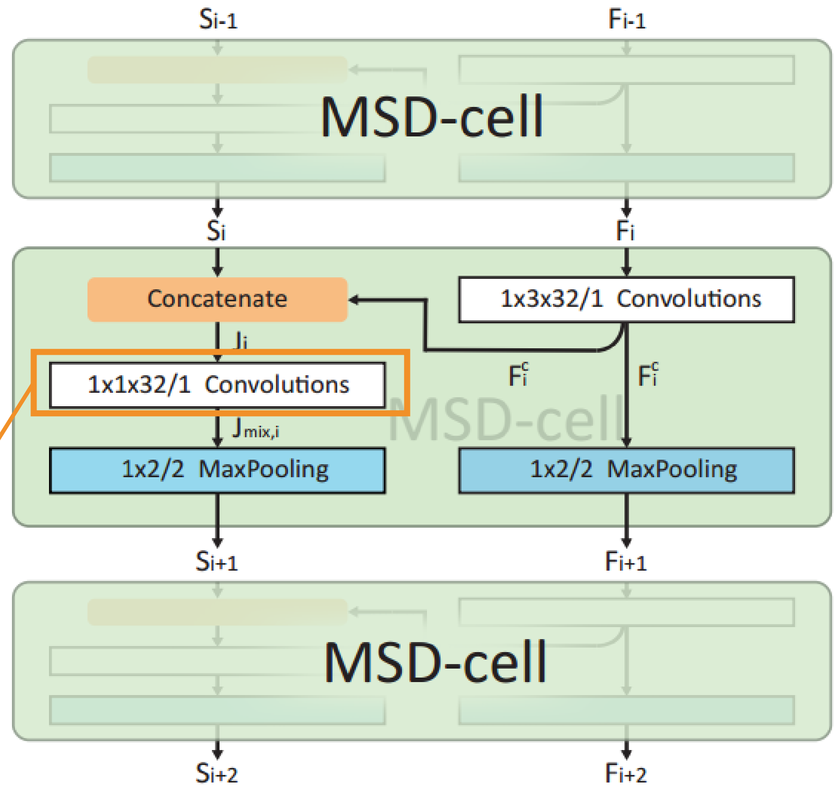


Fig. 3: The detailed structure of MSD-cell, which can be expanded easily.

Methodology

◆ MSD-cell: Generating Multi-Scale Description

the module needs to implement two key functions

First function: extract multi-scale description

- The first function can remember prior features on different scales and add new scale feature.
- Implemented by a 1x3x32/1 convolutions

Second function: mix multi-scale description

- the second function can compare and mix these two kinds of features.
- Implemented by a 1x1x32/1 convolutions

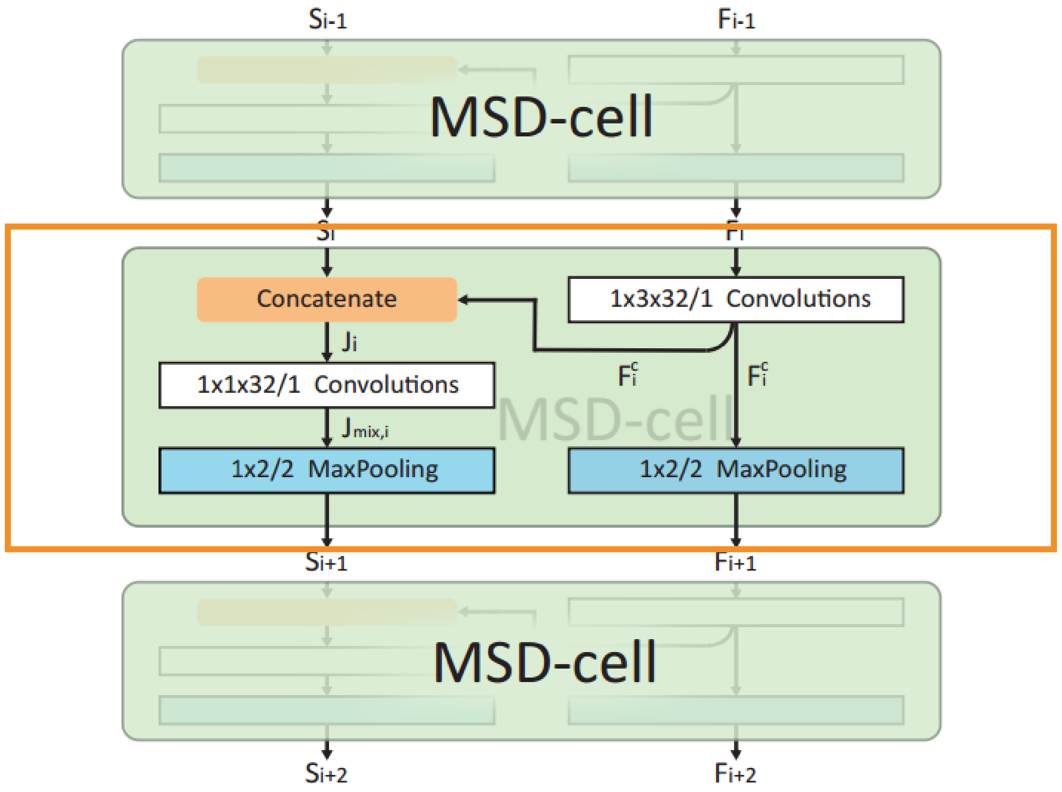


Fig. 3: The detailed structure of MSD-cell, which can be expanded easily.

Methodology

- **MSDNN**

Multi-Scale Description based Neural Network

Shared part

- The shared part can extract multi-scale description feature from seismic waveforms with three spatial dimensions.
- Implemented by 1 conv(1x3) and 10 MSD-cell

Detection part

- the Detection part can distinguish multi-scale description features for earthquake detection.
- Implemented by fc(128), fc(2) and softmax

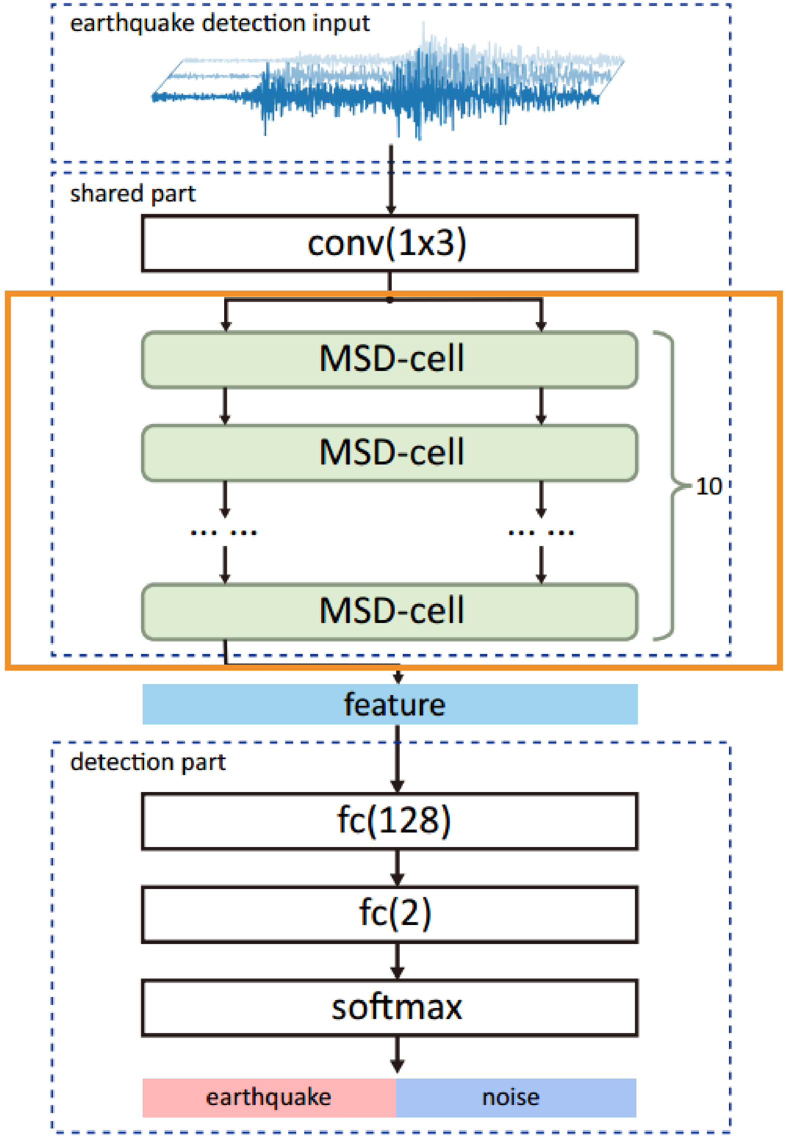


Fig. 4: The framework of MSDNN, which is divided into shared part and detection part.

Methodology

- **MSDNN**

  Multi-Scale Description based Neural Network

  Shared part

  - The shared part can extract multi-scale description feature from seismic waveforms with three spatial dimensions.
  - Implemented by 1 conv(1x3) and 10 MSD-cell

  Detection part

  - the Detection part can distinguish multi-scale description features for earthquake detection.
  - Implemented by fc(128), fc(2) and softmax

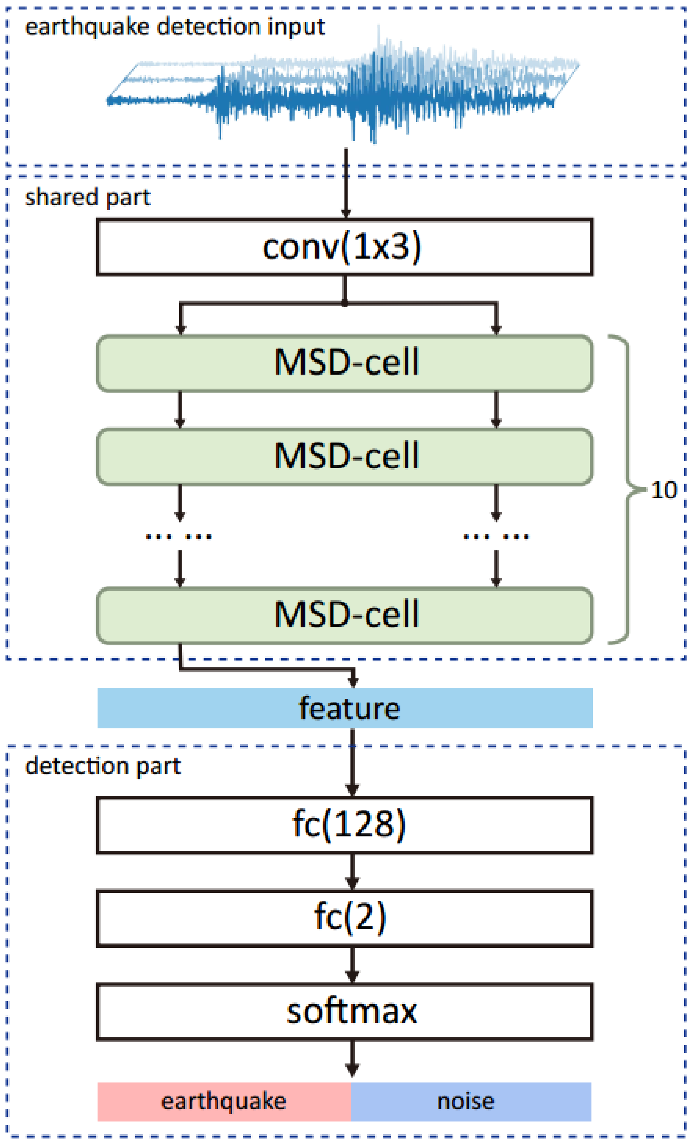


Fig. 4: The framework of MSDNN, which is divided into shared part and detection part.

Homologous Earthquake Waveforms

- Homologous Earthquake waveforms are one earthquake detected by multiple monitoring stations.
- Each pair of waveforms can be label as positive pair (means homology) or negative pair (means non-homology)

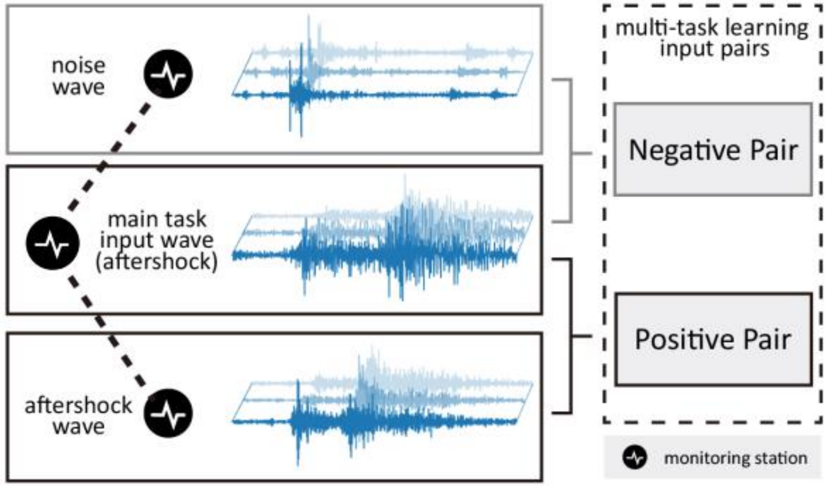


Fig. 5: Sampling pairs of multi-task learning.

Methodology

◆ Homologous Earthquake Waveforms

- Homologous Earthquake waveforms are one earthquake detected by multiple monitoring stations.
- Each pair of waveforms can be label as positive pair (means homology) or negative pair (means non-homology)

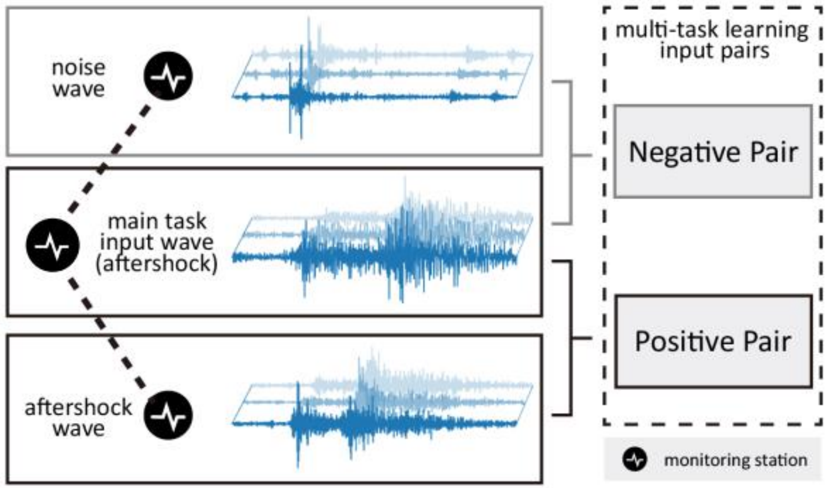


Fig. 5: Sampling pairs of multi-task learning.

Methodology

Homologous Earthquake Waveforms

- Homologous Earthquake waveforms are one earthquake detected by multiple monitoring stations.
- Each pair of waveforms can be label as positive pair (means homology) or negative pair (means non-homology)

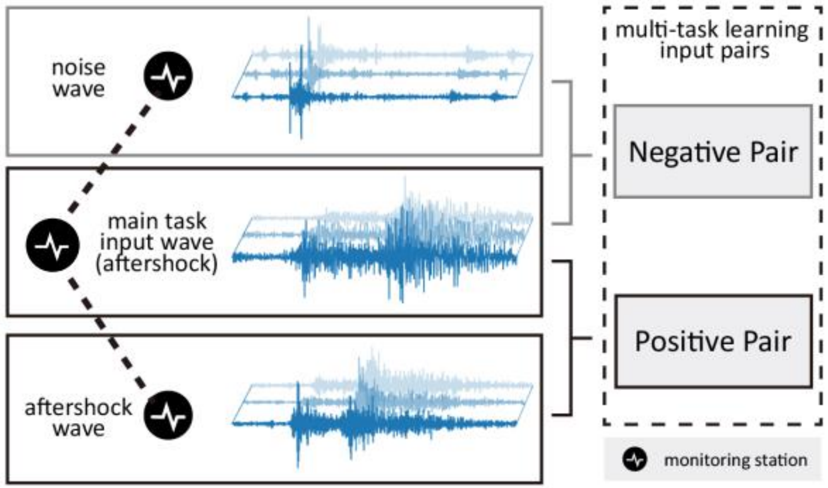


Fig. 5: Sampling pairs of multi-task learning.

Methodology

- **Multi-Task Learning Strategy**

  - **Auxiliary task** : homologous earthquake detection task.

  - **Main task**: aftershock detection task.

  - Auxiliary task can optimize the multi-scale description feature

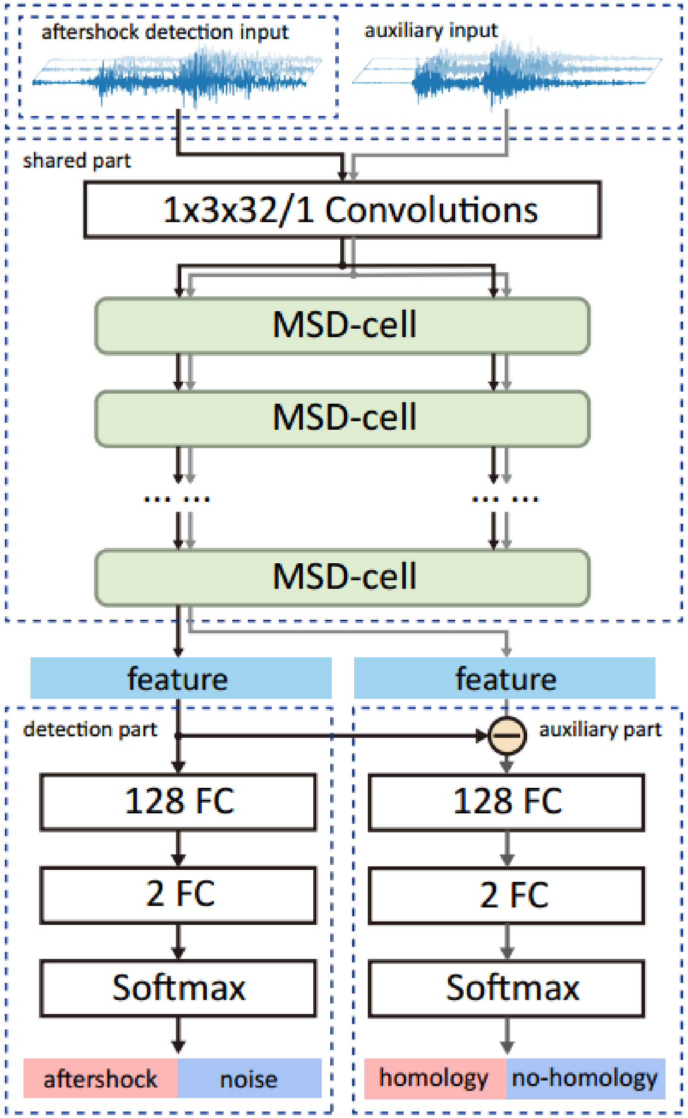


Fig. 4: The framework of MSDNN and multi-task learning, which is divided into shared, detection and auxiliary part.

Methodology

- **Multi-Task Learning Strategy**

  - **Auxiliary task** : homologous earthquake detection task.

  - **Main task**: aftershock detection task.

  - Auxiliary task can optimize the multi-scale description feature

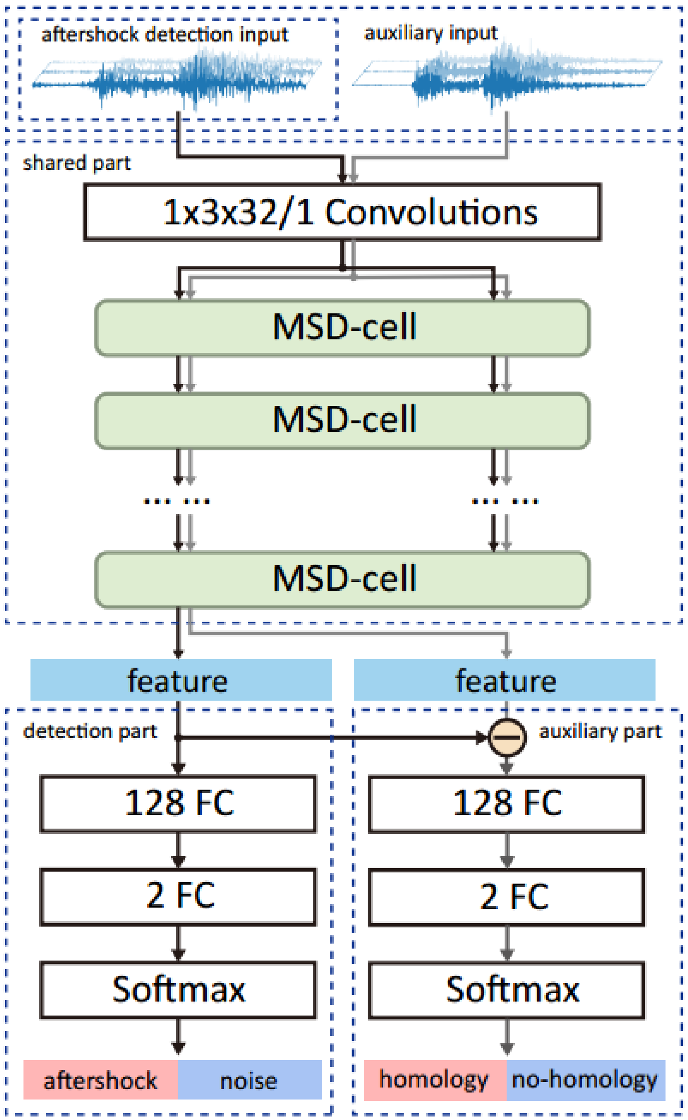


Fig. 4: The framework of MSDNN and multi-task learning, which is divided into shared, detection and auxiliary part.

Methodology

- **Multi-Task Learning Strategy**
  - **Auxiliary task** : homologous earthquake detection task.
  - **Main task**: aftershock detection task.
  - Auxiliary task can optimize the multi-scale description feature

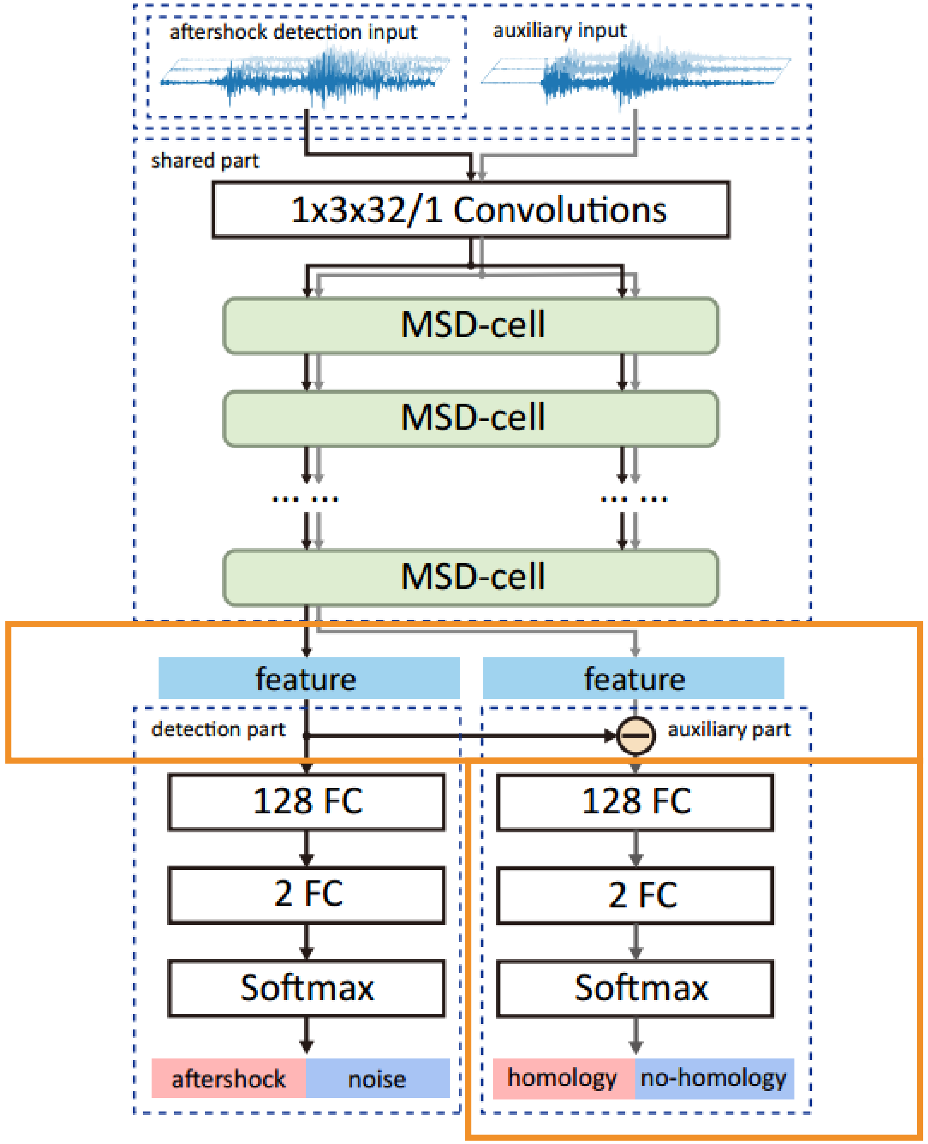


Fig. 4: The framework of MSDNN and multi-task learning, which is divided into shared, detection and auxiliary part.

Experiments

Overall Performance

- our **MSDNN** methods consistently outperform all the baselines.
- the **Multi-task Learning** can improve performance in all metrics.

TABLE III: The overall performance.

Method	Accuracy	Recall	Precision	F_1
Logistic Regression	0.505	0.520	0.080	0.130
Support Vector Machine	0.515	0.520	0.080	0.130
Random Forest [40]	0.767	0.680	0.190	0.300
XGboost [41]	0.882	0.770	0.350	0.490
ConvNetQuake [7]	0.935	0.602	0.544	0.571
Inception Net [11]	0.941	0.637	0.582	0.608
Our Solutions				
MSDNN	0.952	0.638	0.678	0.658
MSDNN+Multi-task Learning	0.954	0.667	0.683	0.675

Experiments

Overall Performance

- our **MSDNN** methods consistently outperform all the baselines.
- the **Multi-task Learning** can improve performance in all metrics.

TABLE III: The overall performance.

Method	Accuracy	Recall	Precision	F_1
Logistic Regression	0.505	0.520	0.080	0.130
Support Vector Machine	0.515	0.520	0.080	0.130
Random Forest [40]	0.767	0.680	0.190	0.300
XGboost [41]	0.882	0.770	0.350	0.490
ConvNetQuake [7]	0.935	0.602	0.544	0.571
Inception Net [11]	0.941	0.637	0.582	0.608
Our Solutions				
MSDNN	0.952	0.638	0.678	0.658
MSDNN+Multi-task Learning	0.954	0.667	0.683	0.675

Experiments

Overall Performance

- our **MSDNN** methods consistently outperform all the baselines.
- the **Multi-task Learning** can improve performance in all metrics.

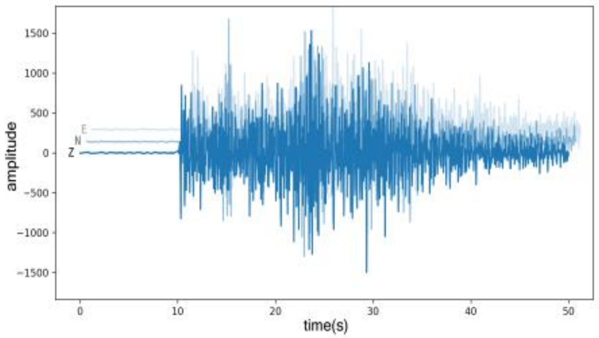
TABLE III: The overall performance.

Method	Accuracy	Recall	Precision	F_1
Logistic Regression	0.505	0.520	0.080	0.130
Support Vector Machine	0.515	0.520	0.080	0.130
Random Forest [40]	0.767	0.680	0.190	0.300
XGboost [41]	0.882	0.770	0.350	0.490
ConvNetQuake [7]	0.935	0.602	0.544	0.571
Inception Net [11]	0.941	0.637	0.582	0.608
Our Solutions				
MSDNN	0.952	0.638	0.678	0.658
MSDNN+Multi-task Learning	0.954	0.667	0.683	0.675

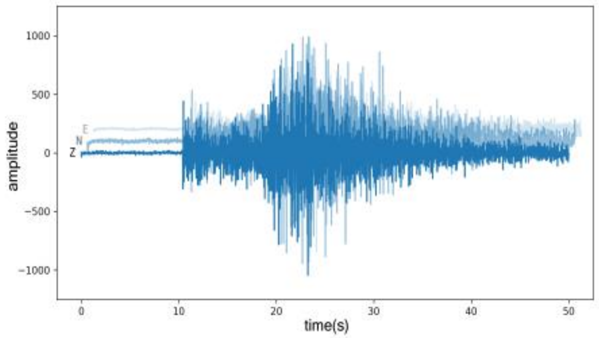
Experiments

Case Studies

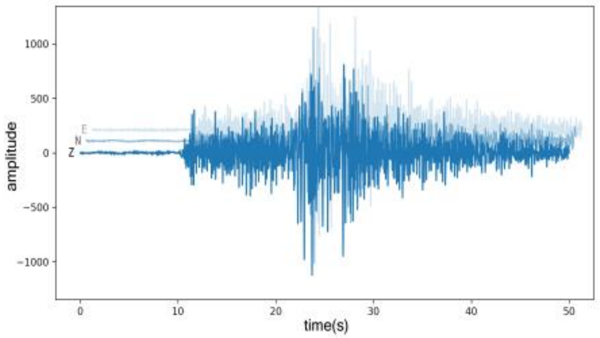
- The “False Positive” waveforms and the “True Positive” waveforms are similar.
- Several geophysical experts re-check 163 “False Positive” waveforms, and 161 waveforms of them were labeled as “Positive”. MSDNN framework can help manual checking.



(a) Average of False Positive



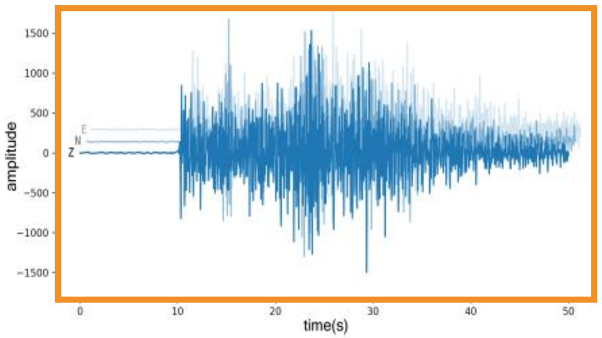
(b) True Positive



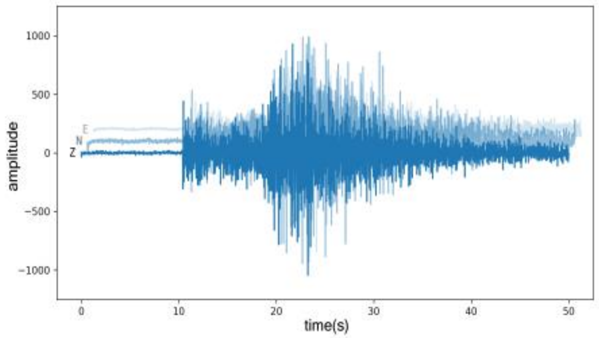
(c) False Positive

Experiments

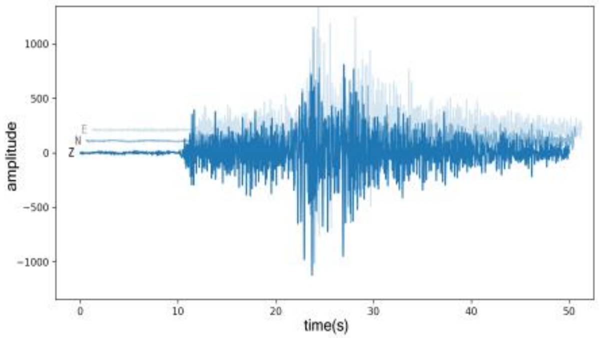
- ## Case Studies
  - The “False Positive” waveforms and the “True Positive” waveforms are similar.
  - Several geophysical experts re-check 163 “False Positive” waveforms, and 161 waveforms of them were labeled as “Positive”. MSDNN framework can help manual checking.



(a) Average of False Positive



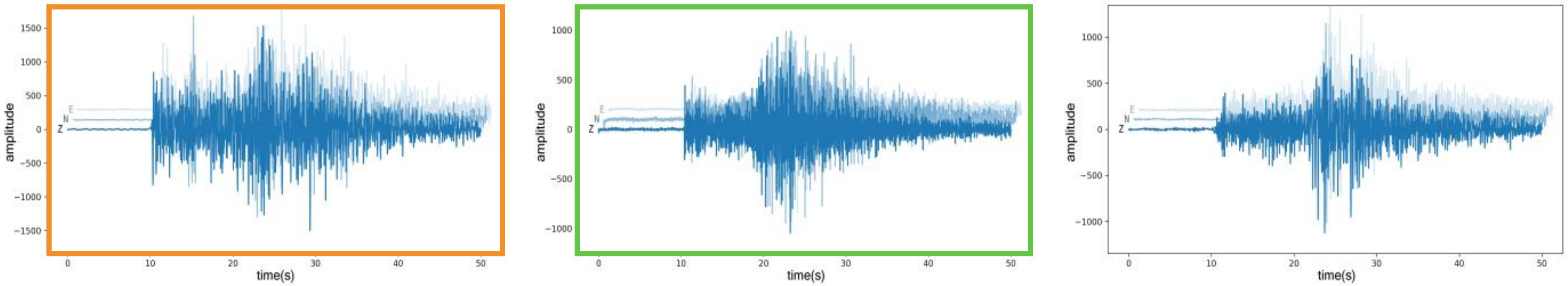
(b) True Positive



(c) False Positive

Experiments

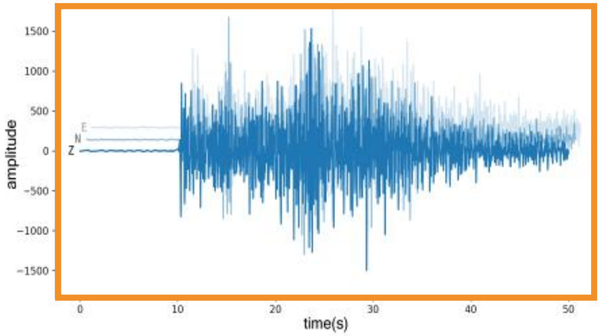
- **Case Studies**
  - The "False Positive" waveforms and the "True Positive" waveforms are similar.
  - Several geophysical experts re-check 163 "False Positive" waveforms, and 161 waveforms of them were labeled as "Positive". MSDNN framework can help manual checking.



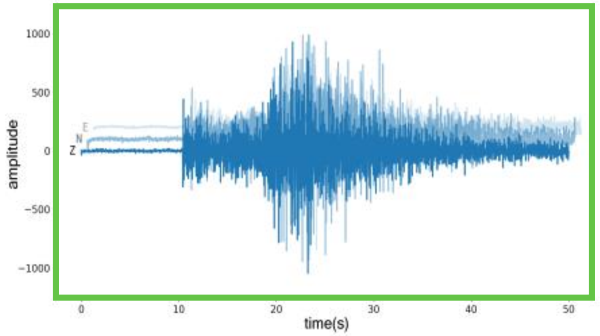
(a) Average of False Positive (b) True Positive (c) False Positive

Experiments

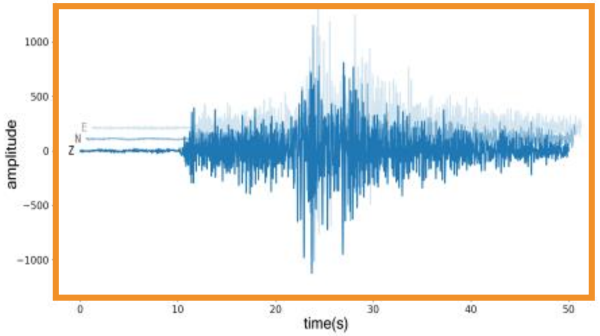
- **Case Studies**
  - The “False Positive” waveforms and the “True Positive” waveforms are similar.
  - Several geophysical experts re-check 163 “False Positive” waveforms, and 161 waveforms of them were labeled as “Positive”. MSDNN framework can help manual checking.



(a) Average of False Positive



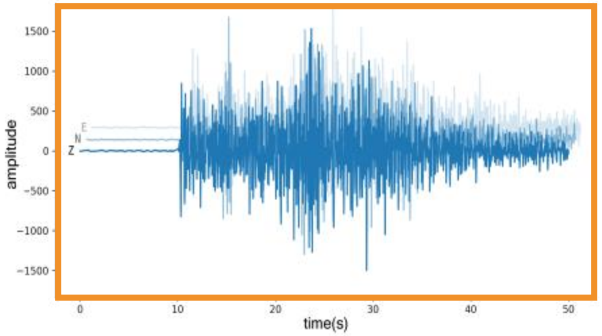
(b) True Positive



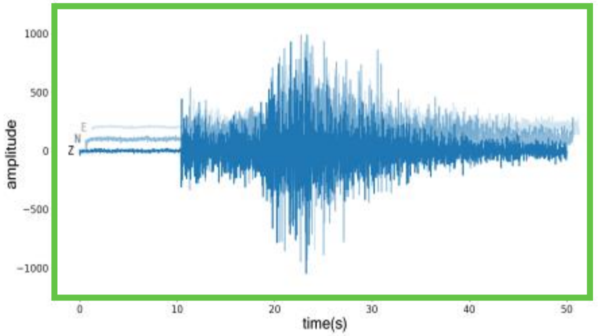
(c) False Positive

Experiments

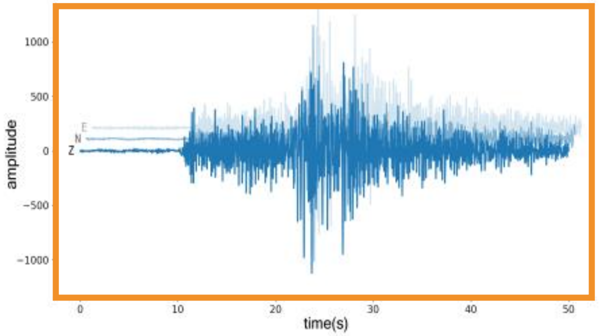
- **Case Studies**
  - The “False Positive” waveforms and the “True Positive” waveforms are similar.
  - Several geophysical experts re-check 163 “False Positive” waveforms, and 161 waveforms of them were labeled as “Positive”. MSDNN framework can help manual checking.



(a) Average of False Positive



(b) True Positive

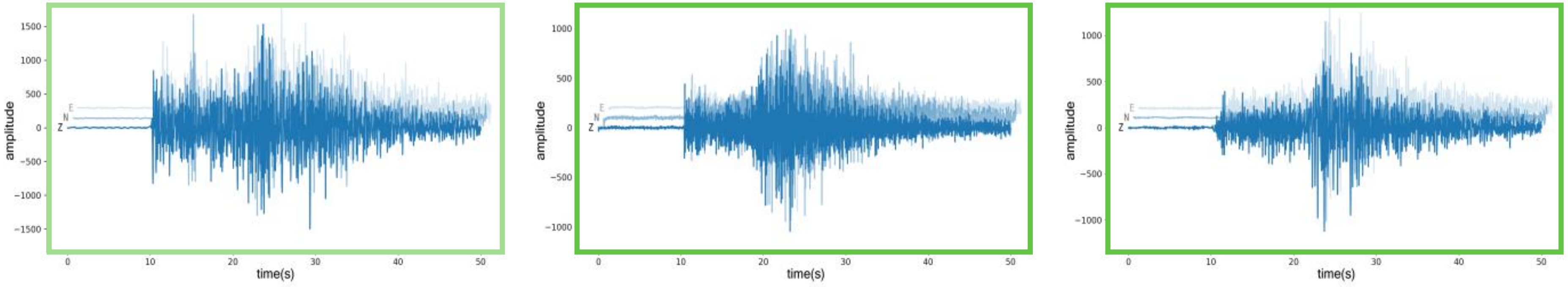


(c) False Positive

Experiments

Case Studies

- The “False Positive” waveforms and the “True Positive” waveforms are similar.
- Several geophysical experts re-check 163 “False Positive” waveforms, and 161 waveforms of them were labeled as “Positive”. MSDNN framework can help manual checking.

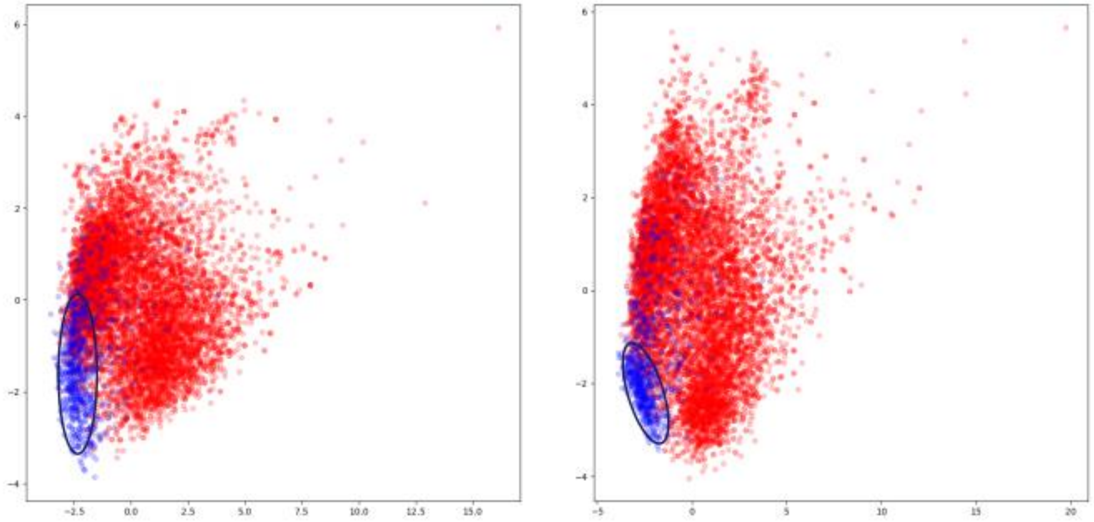


(a) Average of False Positive (b) True Positive (c) False Positive

Experiments

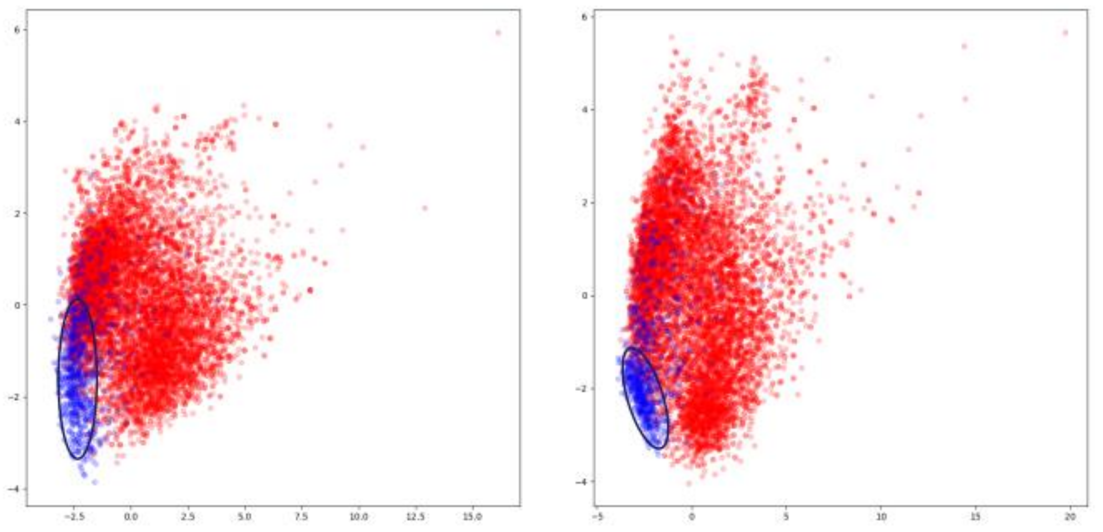
PCA Visualization

- Blue dots are the positive samples and the red dots are negative samples.
- The positive blue dots are close together and easily separated from the red dots when multi-task learning is deployed.



(a) MSDNN (b) MSDNN+Multi-task Learning

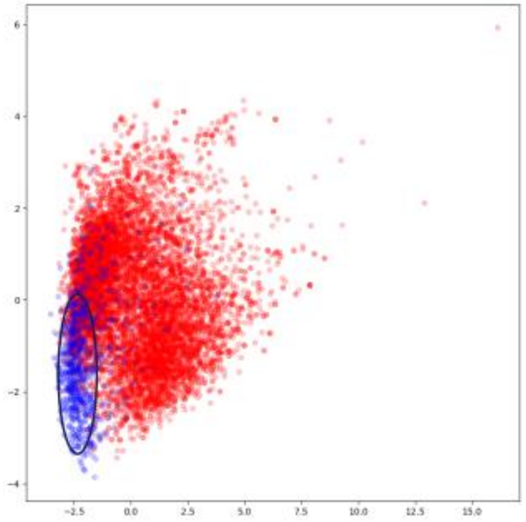
- **PCA Visualization**
    - Blue dots are the positive samples and the red dots are negative samples.
    - The positive blue dots are close together and easily separated from the red dots when multi-task learning is deployed.



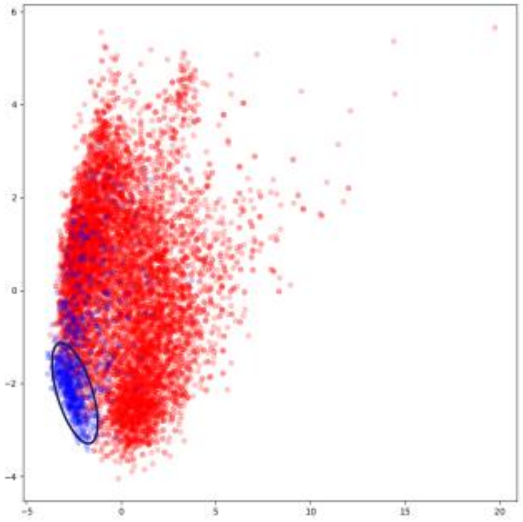
(a) MSDNN (b) MSDNN+Multi-task Learning

Experiments

- **PCA Visualization**
  - Blue dots are the positive samples and the red dots are negative samples.
  - The positive blue dots are close together and easily separated from the red dots when multi-task learning is deployed.



(a) MSDNN



(b) MSDNN+Multi-task Learning

Conclusion

- **Contributions**
  - We propose a novel neural network based solution: **MSDNN (*Multi-Scale Description based Neural Network*)**
  - We design a **multi-task learning** strategy for utilizing the relationship between different monitoring stations.
  - We evaluate our framework on a **real-world data set** from aftershocks of the Wenchuan M8.0 Earthquake.
- **Future Work**
  - Earthquake Rapid Report and Time-series Event Detection

Thank you for listening !
Q&A

Reporter: Qi Zhang

zq26@mail.ustc.edu.cn