This is a peer-reviewed, accepted author manuscript of the following conference paper: Sobot, T, Murray, D, Stankovic, V, Stankovic, L & Shi, P 2024, An active learning framework for microseismic event detection. in 2024 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Piscataway, NJ.

AN ACTIVE LEARNING FRAMEWORK FOR MICROSEISMIC EVENT DETECTION

Tamara Sobot[*], *David Murray*[†], *Vladimir Stankovic, Lina Stankovic*

Electronic and Electrical Engineering
University of Strathclyde
Glasgow, UK

Peidong Shi[‡]

Swiss Seismological Service
ETH Zurich
Switzerland

ABSTRACT

Induced microseismic monitoring has gained increased interest recently, to support various subsurface activities, including geothermal exploration and oil and gas production. To accurately detect and locate origins of microseismisity, deep learning-based methods have become popular due to their high accuracy when trained on large well-labelled datasets. However, though a huge amount of publicly available seismic measurements is available, laballed data to train models is very scarce, since labelling is time consuming and requires very specialist knowledge. Building on our prior work on active learning for time-series data, we propose an active learning method that cleverly picks only a small number of samples to query and stops when the proposed stopping criterion is met. We demonstrate that the proposed approach can save up to 83% of labelling effort even when transferred to a well with different sensing equipment from those used to build the training set.

Index Terms— Active learning, microseismic event detection, deep learning

1. INTRODUCTION

Microseismic monitoring has been gaining attention over the past few years to further illuminate regional-scale induced earthquakes, termed microseismic events, to enable monitoring of subsurface projects, such as oil and gas production, hydraulic fracturing for unconventional resources, e.g., geothermal energy, or the reaction of the Earth's crust to impoundment and storage of water in dams. The injection of fluids into the ground during geothermal energy exploitation fractures the surrounding rock thereby inducing small earthquakes. These microseismic signals are characterised by very low signal to noise ratio (SNR) and hence, unlike earthquakes with relatively higher magnitude and SNR, these signals are challenging to detect in the presence of ambient noise from fluid injection and machinery. With computing resources becoming more and more available, deep neural network-based approaches on data from borehole arrays over the area of underground operations have gained importance in induced microseismic event monitoring, as detailed in recent review paper [1]. These methods are used to provide better insights into underground processes for optimisation of the hydraulic fracturing during injection, as well as for real-time risk evaluation of induced seismicity. Microseismic monitoring starts with detection of microseismic events, which in turns enables localisation of hypocentres, and further characterisation of source mechanisms of microseismic events. The main issue of powerful deep-learning based detection methods, e.g., [2], [3] is that they require a large number of labelled samples, which is time-consuming and requires specialised knowledge. To alleviate this issue, in this paper, we propose an active learning strategy that works in conjunction with a deep learning-based algorithm, to include only most informative samples in the training set and thus reduce the labelling time. This in turn results in improving the speed and consistency of the base detection algorithm, which is a key requirement for microseismic applications [1]. Specifically, we adapt an active learning framework of [4], originally proposed for energy dissagregation, to microseismic event detection which creating high-quality training data sets in a data- and time-efficient way, by labelling only most informative samples. Furthermore, we transfer pre-trained models to new locations and different sensor types (where labelled data unavailable for pre-training), exploiting reliable predictions from multiple sensors at a time to make a final decision. Results show that active learning brings improvement to the detection algorithm performing on both a new location and different sensor type,

[*]This research was partly funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955422.

[†]This work was partly supported by the Engineering and Physical Sciences Research Council New Horizons research programme EP/X01777X/1

[‡]P.S. was supported by the De-Risking Enhanced Geothermal Energy project (DEEPs). DEEP is subsidized through the Cofund GEOTHERMICA (Project No. 200320-4001), which is supported by the European Union's HORIZON 2020 programme and various National Funding Agencies for research, technological development, and demonstration under Grant Agreement Number 731117.

while saving 83% of labelling effort.

2. ACTIVE LEARNING FOR SEISMIC MONITORING

Active learning [5] is a method used to reduce labelling effort of machine learning models. It is based on the assumption that different data samples bring different amount of information to the model if included in training, and therefore, only highly-informative samples can be chosen to be labelled and used in training, without compromising performance.

An extensive review of deep active learning approaches can be found in [6] covering a range of applications. However, active learning, and more generally, data efficient approaches, are rarely studied in the context of microseismic event detection. An automated active learning approach has been proposed in [7] for interpretation of seismic volumes. The method involves a CNN for seismic image processing, for identification of facies. [8] proposes an active learning framework for seismic image interpretation, with acquisition function that combines deep clustering and uncertainty sampling. An active learning method for classification of time-series volcano-seismic events is proposed in [9], and includes a CNN and a diversity-based acquisition function. In contrast to the above approaches that are either based on image data [7], [8], or on times-series that of high magnitudes and have high SNR such in volcano seismisity detection [9], in this paper we propose an active learning framework for time-series microseismic signals that are challenging to detect due to their short duration, low magnitudes, and high noise. The proposed algorithm processes uses optimal thresholding [4] to select samples to be labelled and a criterion to stop once performance cannot be further improve.

3. METHODOLOGY

3.1. Dataset

The dataset used in this paper is recorded as part of the Utah Frontier Observatory for Research in Geothermal Energy (FORGE) project, designed for research on creating, sustaining, and monitoring enhanced geothermal systems (EGS). Hydraulic stimulation was conducted in the target heat reservoir to create fractures and increase the rock permeability in the hot dry rock. Pressurised fluids are injected into the deviated stimulation well 16A-32 at a depth of around 2.4 km at three different stimulation stages. The stimulation at Utah FORGE is monitored via seismic instrumentation deployed in deep boreholes. During the last stimulation, three deep vertical monitoring wells (58-32, 56-32, and 78B-32) distributed in different azimuths were in place to capture the stimulated fractures, i.e., induced microseismic events. Well 58-32 and well 78B-32 contain 8-level 3-component digital geophones (represented as G1 to G8), with each level separated by 100 feet. In well 56-32, however, measurements are acquired using a two-level analog seismic monitoring tool (G1 and G2). In this paper, we analyse one hour of continuous data recorded at a sampling rate of 4 kHz (from 20:00 to 21:00 on 21 April 2022) recorded by the 18 sensors from the three monitoring wells during the third stimulation stage.

Monitored events are of induced microseismic nature, they are pulse-like and have high-frequency components. To remove low-frequency noise from machinery at the site, and also high-frequency measurement noise, obtained signals are filtered using a fourth-order Butterworth band-pass filter, with the passband between from 100 to 1800 Hz. An example of an event from well 58-32, the bottom geophone (G8) is shown in Figure 3. Data is filtered as described above. FP1 and FP2 are horizontal, and FPZ is the vertical seismogram component. Horizontal axis represents time (in samples; 1 sample = 0.25msec), and vertical sensor measurements in mm/sec. Ground truth indicates the time between P and S wave arrivals. An event catalog [10] is generated with the EQ-Transformer of [11], pretrained using a global distribution of earthquakes. It contains timestamps of P and S wave arrivals associated with each event for each sensor where it was picked, together with corresponding signal-to-noise (SNR) values. After automatic label generations with Earthquake Transformer, the predicted labels were validated by seismologists via visual inspection.

3.2. Active learning framework

Active learning relies on cleverly selecting a small number of unlabelled samples to be send to an expert for labelling. A good active learning model reaches high performance with a very small amount of samples sent for labelling. The set of unlabelled data samples from which the informative ones are chosen, is called *query pool*. Active learning works iteratively - in each iteration, samples from the query pool are ranked by the uncertainty of the model when making predictions on them (only samples that model is uncertain about should be labelled), or by a distance from already labelled data samples (only samples that bring new information should be added to the training set), or combination of two approaches. Function that is used to rank and choose samples to label is called *acquisition function*. After the samples are chosen, they are removed from the query pool, sent for labelling and then used to fine-tune the model. Performance is expected to improve quickly in the first iterations of the process, and to slow down later, when the amount of novelty brought by newly labelled samples reduces. The process runs until the stopping criterion is met. A block diagram of active learning framework is presented in Figure 1.

In this paper, the used acquisition function is optimal thresholding based, with stopping criterion as proposed in [4]. Namely, the deep learning model returns a prediction window containing values between 0 and 1, indicating if the

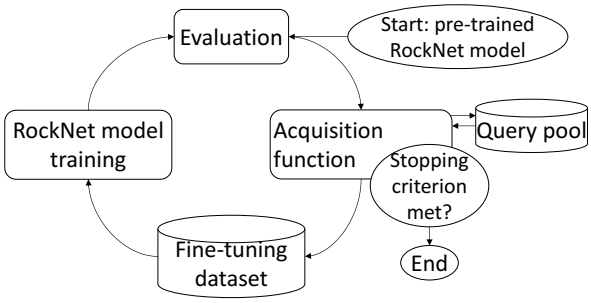

Fig. 1: Active learning framework

event is detected at each timestamp inside the window. Then, the maximum of all prediction values in the window determines a single prediction value for the window. To determine if an event is detected or not, we set a threshold $0 < T < 1$. The interval $(0, 1)$ is split into three regions: a region $[0, T/2)$ where the model is certain that the window does not contain events, i.e., a negative prediction; a region where the model is uncertain about its prediction $[T/2, (1 + T)/2)$; and a region $[(1 + T/2), 1]$ where the model is certain that the window contains an event, i.e., positive prediction. Then, samples belonging to the uncertain prediction region are ranked based on their distance from the decision threshold T, and a number $(batchsize/2)$ of all these uncertain samples that are the closest to T is queried. Additionally, a number of samples $(batchsize/4)$ is queried randomly from predictions that fall into certainly negative and a number $(batchsize/4)$ from certainly positive prediction regions. This constitutes a batch of samples that is queried, labelled and included in fine-tuning dataset. The largest number of samples is chosen from the uncertain region because those samples are supposed to bring the most information to the model. However, to prevent the model from overfitting and forgetting patterns of positive and negative samples for which it is typically certain about, samples are chosen from certain predictions as well. Active learning stops when there are three consecutive epochs with empty uncertain region.

To demonstrate efficiency of the proposed active learning process, we use a popular RockNet network [3]. RockNet is a fusion model, taking both 3-channel time series window, and a spectrogram of the vertical channel of the window as inputs, and making a fusion of features extracted from both inputs. The model returns masks showing detected events - earthquakes and rockfalls. For earthquakes, masks are positive between P and S wave arrivals, and negative elsewhere. The model demonstrates excellent performance for detecting earthquake and rockfall events [3]. In this paper, we apply it to detect induced microseismic events, of much shorter duration and lower amplitudes, and also higher noise levels.

3.3. Experimental setup

Recordings are divided into two periods - the last 15 minutes is reserved for testing, while the first 45 minutes are used as training set if pre-training, or query pool if performing active learning. This 45-15 min split is the same for all sensors. Since for induced microseismic events time between P and S phase arrivals is very short (less than 100 ms), the input window length that is used in this paper is 400 ms, or $W = 1600$ samples. Spectrograms are generated from vertical seismogram component of the input window, with the FFT length $N = 128$, and a step of $H = 10$ samples between STFT segments. So, the input dimensions are (3, 1600) for time-series input and (2, $N/2 + 1 = 65$, $W/H + 1 = 161$) for the spectral domain input - one channel for real and one for imaginary spectrogram component. Since there is only one type of event in the dataset, at the output we only have one binary mask, of dimensions (1, 1600), which is positive between P and S wave arrivals, and negative elsewhere. Training windows overlap by $1/3$, while testing windows do not overlap. For each sensor, there is 24,779 samples used as training set if the sensor is used for pre-training, or as query pool if active learning has been applied to the sensor; and 2253 test samples. In the pre-training phase, training data was balanced by keeping an equal number of windows that do and do not contain an event. Excessive windows with no events present are discarded. Testing data was not balanced to reflect real-world recordings. Query pool data was also not balanced.

Three experiments were conducted in this research, to demonstrate value of proposed active learning framework across different levels of generalisation to exploit multiple sensors in the borehole arrays on the site: (1) The deep learning model is pre-trained with data from two sensors from the well 78B-32 (sensors G7 and G8), and active learning is performed with another sensor data from the same well (G5), to verify the active learning approach. (2) The deep learning model is pre-trained with data from four sensors from the well 78B-32 (G5-8), and active learning is performed with the data from a sensor from the well 58-32 (G5), to test generalisation across different wells with the same instrument types - the two wells are distributed in different azimuths which can impact the signals. (3) The deep learning model is pre-trained with the data from 8 sensors from the wells 78B-32 (G5-8) and 58-32 (G5-8), and active learning is performed with the data from a sensor from the well 56 (G2). This is to test generalisation across different wells and different instrument types, since well 56-32 has different instrumentation than the other two wells.

Model performance is evaluated using F_1-score: $F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} = \frac{TP}{TP + \frac{1}{2} \cdot (FP + FN)}$, where TP denotes true positives, FP false positives, and FN false negatives.

The stopping criteria as defined in [4], is based on the optimal point of active learning, defined as the point on the active learning curve closest to the "ideal" point $(0, 1)$ - no data samples labelled, $F1$-score equal to 1.

Batch size used is 128, learning rate $1e - 3$, and decision threshold $T = 0.5$ for Experiments 1 and 3, and $T = 0.6$ for Experiment 2. These model thresholds are heuristically

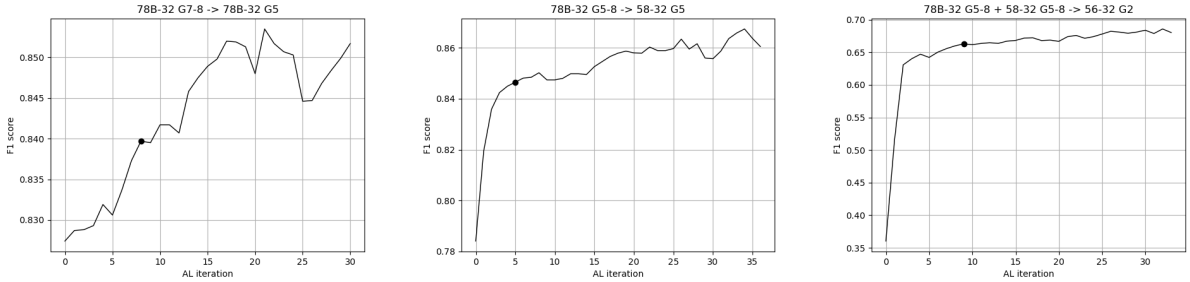

Fig. 2: Results: Experiment 1 - left; Experiment 2 - middle; Experiment 3 - right.

set based on the testing data during the pre-training phase and are not further tuned during the active learning phase. In the pre-training phase, training is performed for a maximum of 50 epochs with early stopping patience of 5. In the active learning phase, each training is performed for 15 epochs, and the best model is used to make queries.

4. RESULTS & DISCUSSION

The results are presented in Figure 2 for the three experiments discussed in Subsection 3.3. Active learning iterations are presented on the horizontal axis, while the vertical axis shows the resulting F_1-score. Optimal points are marked by dots.

In Experiment 1, where active learning is used in the same well as the pre-training data comes from, active learning curve rises the slowest - the data comes from the same source and is of similar quality as pre-training data, and there was enough pre-training data for the model to learn well. However, the performance does improve compared to the baseline model, and with significantly less data than if the whole query pool was labelled - only 15.5% of query pool is labelled (3840 out of 24779 samples, labelled over 30 iterations), saving 84.5% of labelling time. This indicates that microseismic event detection can benefit from active learning to reduce labelling effort and improve performance.

In Experiment 2, which transfers a pre-trained model to a different well with the same type of sensors, active learning gives promising results - $F1$-score is improved from 0.78 before any fine-tuning to 0.86 at the and of fine-tuning with active learning, with only 18.6% of query pool labelled, reducing labelling effort by 81.4%. This scenario benefits more from active learning because the new data used for fine-tuning during the active learning phase is more informative due to environmental differences in two wells.

In Experiment 3, which transfers a pre-trained model to a different well with different measuring equipment, performance is the most improved. It was poor at the beginning ($F_1 = 0.36$), but with active learning it is quickly improved after only a couple of iterations, reaching $F_1 = 0.68$ with 17% of data samples labelled, reducing labelling effort by 83%. This is due to the fact that newly introduced data is very different from the data used for pre-training, and active learning provides significantly new information.

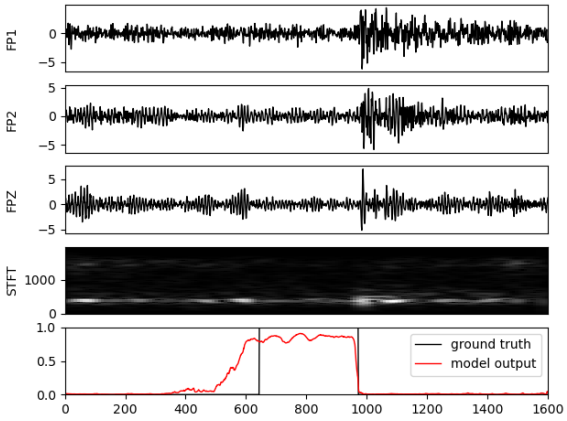

Fig. 3: Model performance example - time-series input (FP1, FP2, FPZ), spectral input (STFT map), and ground truth and prediction for a data sample from sensor G8, well 58-32, Experiment 2.

An example of model performance on a data sample from sensor G8 from well 58-32 in Experiment 2 is shown in Figure 3. The model has 92696 trainable parameters, and processes on average 12.7 batches (12.7*128=1625.6 samples) per second in evaluation mode on a PC with 32GB RAM and an NVIDIA Titan Xp graphic card.

5. CONCLUSIONS

We demonstrated efficiency of active learning for detection of induced micro-seismic events. We showed that by using the proposed approach to selecting most informative samples to be labelled, the labelling effort can be reduced by 80-84 % without affecting the performance. The proposed method is very efficient when transferred to a very different environment with different measuring equipment. It would be worth exploring active learning with clustering approaches in future work, as well as usage of explanation tools which could inform the expert about the reasoning behind model decision and aid labelling.

6. REFERENCES

[1] Denis Anikiev, Claire Birnie, Umair bin Waheed, Tariq Alkhalifah, Chen Gu, Dirk J Verschuur, and Leo Eisner, "Machine learning in microseismic monitoring," *Earth-Science Reviews*, p. 104371, 2023.

[2] Jiaxin Jiang, Vladimir Stankovic, Lina Stankovic, Emmanouil Parastatidis, and Stella Pytharouli, "Microseismic event classification with time-, frequency-, and wavelet-domain convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[3] Wu-Yu Liao, En-Jui Lee, Chung-Ching Wang, Po Chen, Floriane Provost, Clément Hibert, Jean-Philippe Malet, Chung-Ray Chu, and Guan-Wei Lin, "Rocknet: Rockfall and earthquake detection and association via multi-task learning and transfer learning," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[4] Tamara Todic, Vladimir Stankovic, and Lina Stankovic, "An active learning framework for the low-frequency non-intrusive load monitoring problem," *Applied Energy*, vol. 341, pp. 121078, 2023.

[5] Burr Settles, "Active learning literature survey," 2009.

[6] Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B Gupta, Xiaojiang Chen, and Xin Wang, "A survey of deep active learning," *ACM computing surveys (CSUR)*, vol. 54, no. 9, pp. 1–40, 2021.

[7] Haibin Di and Aria Abubakar, "Automated active learning in seismic image interpretation," *The Leading Edge*, vol. 41, no. 9, pp. 628–635, 2022.

[8] Xiaofeng Gu, Wenkai Lu, Yile Ao, Yinshuo Li, and Cao Song, "Seismic stratigraphic interpretation based on deep active learning," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[9] Grace F Manley, Tamsin A Mather, David M Pyle, David A Clifton, Mel Rodgers, Glenn Thompson, and John Makario Londono, "A deep active learning approach to the automatic classification of volcano-seismic events," *Frontiers in Earth Science*, vol. 10, pp. 807926, 2022.

[10] Peidong Shi, Francesco Grigoli, Federica Lanza, Gregory C Beroza, Luca Scarabello, and Stefan Wiemer, "Malmi: An automated earthquake detection and location workflow based on machine learning and waveform migration," *Seismological Society of America*, vol. 93, no. 5, pp. 2467–2483, 2022.

[11] S Mostafa Mousavi, William L Ellsworth, Weiqiang Zhu, Lindsay Y Chuang, and Gregory C Beroza, "Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking," *Nature communications*, vol. 11, no. 1, pp. 3952, 2020.