Automated picking of seismic first arrivals using single- to multi-domain self-trained network

Mitsuyuki Ozawa[1]

[1]JGI, Inc.,

November 26, 2022

Abstract

Over the last decade, the evolution of survey acquisition design composed of numerous receivers and shot points has exponentially increased the seismic data volume. As a result, the demand for efficient algorithms for detecting seismic first arrivals has been growing. Machine learning methods, especially convolutional neural networks, have shown great potential in previous studies. However, most existing methods lack generalization ability or be sensitive to noise. This paper proposes a new picking strategy based on the single-to multi-domain selftrained network to achieve automated picking for the first arrivals. Our self-trained network learns from a single domain to multiple domains while rejecting the wrong predicted picking and assigning the pseudo-first-arrivals-times to the unlabeled dataset. The model inherits the advantage from both single-and multi-domain picking strategies; it has high generalizability and noise robustness. In addition, the model is designed for a limited amount of manually annotated data to reduce human labor and total seismic processing time. In experiments on an open hard rock seismic multi-survey dataset, our picking method outperforms the benchmark on most evaluation metrics. Remarkably, our approach achieves approximately 10 % accuracy improvement against Lalor site data with a sampling rate that differs from training data. The experimental results demonstrate our method's high generalizability and robustness for unseen datasets.

Automated picking of seismic first arrivals using single- to multi-domain self-trained network

Mitsuyuki Ozawa

JGI, Inc.,1-5-21, Otsuka, Bunkyo-ku, Tokyo 112-0012, Japan

Corresponding Author

Mitsuyuki Ozawa mitsuyuki.ozawa@jgi.co.jp

ABSTRACT

Over the last decade, the evolution of survey acquisition design composed of numerous receivers and shot points has exponentially increased the seismic data volume. As a result, the demand for efficient algorithms for detecting seismic first arrivals has been growing. Machine learning methods, especially convolutional neural networks, have shown great potential in previous studies. However, most existing methods lack generalization ability or be sensitive to noise. This paper proposes a new picking strategy based on the single- to multi-domain self-trained network to achieve automated picking for the first arrivals. Our self-trained network learns from a single domain to multiple domains while rejecting the wrong predicted picking and assigning the pseudo-first-arrivals-times to the unlabeled dataset. The model inherits the advantage from both single- and multi-domain picking strategies; it has high generalizability and noise robustness. In addition, the model is designed for a limited amount of manually annotated data to reduce human labor and total seismic processing time. In experiments on an open hard rock seismic multi-survey dataset, our picking method outperforms the benchmark on most evaluation metrics. Remarkably, our approach achieves approximately 10 % accuracy improvement against Lalor site data with a sampling rate that differs from training data. The experimental results demonstrate our method's high generalizability and robustness for unseen datasets.

INTRODUCTION

In seismic processing, first-break picking, determining the onsets of first signal arrivals, is a fundamental step in estimating subsurface velocity structure through static

correction(Palmer, 1980). Traditionally, picking was carried out by a human visual inspection of the variation of amplitudes and waveforms. However, more advanced seismic acquisition and processing systems have been developed over the past decades. Notably, more advanced survey design composed of numerous receivers and shot points, like 3D seismic survey, has usually been adopted to improve the accuracy and resolution of seismic image. Consequently, manual phase picking of such inflated seismic data volume could be time-consuming and a bottleneck of the data processing flow. Hence, there is a demand for an automated first-break picking technique to reduce human labor. Therefore, many efforts have been made to develop the automatic picking method in the previous decade.

One of the automated picking strategies focuses on the particular feature of the seismic signals, such as the Short-Time Average and Long-Time average ratio(STA/LTA) (Allen, 1978; Stevenson, 1976), Energy Ratio(Coppen, 1985, Han et al., 2009), correlation properties (Peraldi & Clement, 1972), fractal dimension (Boschetti et al., 1996), Akaike information criterion (Takanami & Kitagawa, 1991; Diehl et al., 2009), high-order statics (Yung and Ikellee, 1997) to identify the mutation points. These traditional picking methods are classified into two categories by their operation domain: (1) single-domain and (2) multi-domain (or array processing)-based algorithms (Akram and Eaton (2016)). Single-domain methods only use a signal from an individual receiver trace and do not use adjacent receiver information. In contrast, multi-domain methods, such as cross-correlation-based algorithms, make simultaneous use of information on multiple receiver levels within the array. Because single-domain methods lose the spatial information of the receiver array, they tend to work well only under an adequate signal-to-noise ratio. As a result, they often get unsatisfactory results for low-quality data. In addition, although

the multi-domain methods mitigate the noise effects to some degree, it still struggles with poor data quality.

Another automated strategy is a machine learning based on neural networks. The networks automatically extract and transform sensitive attributes or features of the onset of the seismic signal with the manually annotated label. This strategy appeared in the 1990s, and some authors applied neural networks for estimating first arrivals (Murat & Rudman, 1992; McCormack et al., 1993). Most initial trials used neural networks are based on fully connected (FC) layers where all the inputs from one layer connect to every activation unit of the next layer. Because FC Layers hardly analyze the spatial structure of inputs, they often fail to capture a crucial feature of first arrivals. More recently, many deep learning applications for arrivals picking have been developed (e.g., Ozawa et al., 2018; Yuan et al., 2018; Zhu & Beroza, 2019). Most of those methods are based on the convolutional neural network architecture(CNN), an extended neural network imitating the visual perception mechanism of the living creature (LeCun et al., 1989). They have achieved high performance in various computer vision tasks (e.g., Krizhevsky et al., 2017; Girshick et al., 2014). CNN explicitly assumes that inputs have a specific series structure, like images. This assumption allows us to efficiently encode crucial time-space features of first arrivals from the input signals. Zhu and Beroza (2018) proposed a single-domain picking method for a single three-component record based on a U-shaped fully convolutional network (U-net) (Ronneberger et al., 2015) to estimate the onset of P-wave and S-wave. Yuan et al. (2018) applied CNN to classify whether a local multitrace patch contains first arrivals. Hu et al. (2019) used U-net to recognize the inherent patterns of first arrivals on multi-domain. They directly apply the trained U-net model to another seismic dataset. Their direct transfer test shows that the U-net model can estimate the first arrivals on a simple dataset without

retraining the network because both datasets have a similar first break pattern. However, if the test data set has different first arrivals patterns than the training dataset, U-net should be retrained with parts of the test data.

As mentioned before, the multi-domain methods have obvious advantages in terms of robustness against noise compared with the single-domain methods. On the other hand, in terms of supervised machine learning, single-domain learning has some advantages in the model generalization ability. The diversity of seismic acquisition parameters, such as receiver interval and sampling rate, and a limited amount of supervised data often result in overfitting in neural networks. This issue tends to become apparent in multi-domain supervised model learning because the multi-domain model learns the spatial information of a specific training data field. In contrast, because single-domain learning only uses a single trace, it can learn relatively similar first arrival patterns of the test dataset from the training dataset. In addition, we can provide a relatively large amount of training datasets for single-domain learning. These advantages increase the single-domain model's generalizability for unseen data. Loginov et al. (2022) compared several neural network architectures, including single-domain and multi-domain approaches for first arrivals picking for land seismic data. Their study shows the single-domain approach that provides the most reliable results in the case of first-arrival travel-time variations typical of complicated near-surface structures.

Self-training (Yarowsky, 1995) is one of the approaches in semi-supervised learning (Fralick, 1967), which train a model to learn pseudo-labels predicted by another previously-learned model. The algorithm starts by learning a supervised classifier on the labeled training set. Then, the classifier selects the part of the unlabeled dataset using the confidence of the

classifier's prediction and assigns pseudo-labels to them. The pseudo-labeled examples are then used to enrich the training data, and the process is repeated until the algorithm converges.

This study proposes a new strategy based on a single- to multi-domain self-trained network to improve automated first arrivals picking accuracy and stability. Our self-trained network iteratively learns while rejecting the poor pick and assigning the pseudo-first arrivals times to the unlabeled dataset. One characteristic of our method is it starts from single-domain training and then shifts the multi-domain training. Hence, our model is a hybrid of single-domain and multi-domain convolution-based networks and inherits the advantage of both domain strategies; the generalized ability of the single-domain network and the noise robustness of the multi-domain network. In addition, the model is designed for a limited amount of manually annotated data, which can reduce human labor and total seismic processing time.

Some proposed methods have already used the preliminary predicted pick or extracted attribute to improve the final picking results. Sabbiones and Velis (2010) proposed a poor trace-by-trace picks correcting technique to exploit the benefits of the data present in the entire shot record. Wang et al. (2022) used the U-Net as the coarse segmentation network to obtain the approximate first positions of break arrivals. Then, they pick up the data and predict the first arrivals using a refine-segmentation network based on U-Net. Yuan et al. (2020) used U-Net to segment seismic images and recurrent neural networks for arrival picking. Duan & Zhang (2020) applied the short- and long-time window average ratio (STA/LTA) method to obtain the preliminary picking and then fix poor picks based on the multitrace analysis using CNN. Tsai et al. (2018) introduced deep semi-supervised machine learning for first arrivals pickings. They extracted the significant features from unlabeled data using an unsupervised autoencoder to prepare the input of supervised training. Chen et al. (2019) proposed combined k-means

clustering with CNN classification for arrival picking. Their CNN classified the first waveform from the entire signal, and then k-means clustering picked the first arrivals from the first waveform. The differences between these methods and our proposed self-trained method are below. (1) Our method directly uses preliminary picking as supervised data. Therefore, our strategy has the potential to extract more information from the preliminary pick to improve prediction accuracy. (2) Our neural network learns from single- to multi-domain in order to improve both generalized ability and noise robustness.

The remainder of this paper is organized as follows. The next section presents the new workflow based on a self-training deep learning model. In section 3, we evaluate our model using an open seismic multi-survey dataset. We discuss the capability and limitation of our model and future prospects in section 4

Method

The workflow of our single- to multi-domain self-trained learning method consists of three stages (Figure 1);

1) Single-domain network learning using the manually labeled dataset.

2) Predicting first arrivals for an unlabeled dataset using a trained network and automatically selecting the highly confident prediction to assign pseudo-labels.

3) Multi-domain network learning using the assigned pseudo-labels and predicting first arrivals for an unlabeled dataset.

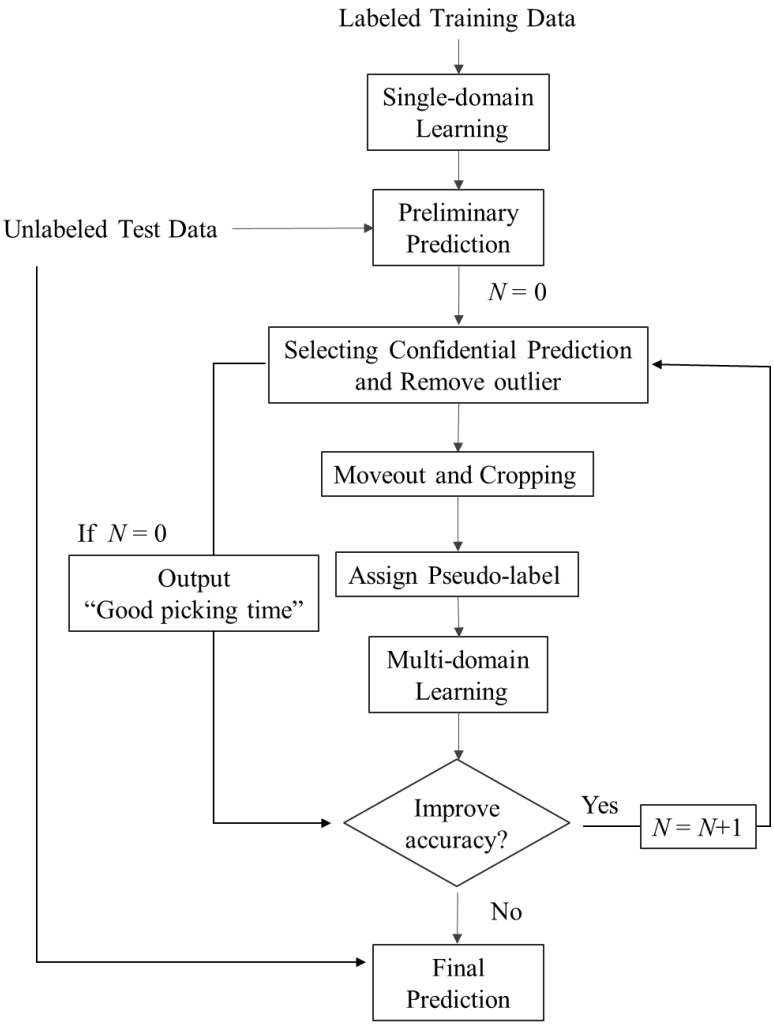

Figure 1. Workflow for first-break piking using the single to-multi-domain self-trained network.

Single-domain network learning.

Our workflow starts from single-domain learning using the manually labeled training data as supervised data. Before single-domain learning, we normalize each training waveform by removing its mean and dividing it by the standard deviation. The manual first arrival times are converted to binary labels with 1 for pixels labeled as "first arrivals" and 0 otherwise. As a single-domain learning network, we use the 1D U-net architecture, a modification of U-net, to deal with 1-D time series data. The input is a single trace of the dataset. The network output and

the ground truth label (binary label) are compared using a root mean squared error loss function. The network is trained to minimize the loss using the adaptive moment estimation (Adam) optimizer (Kingma & Ba, 2014).

Generating pseudo-labels

After training a single-domain network, we apply the trained model to the unlabeled dataset. The self-training process needs to determine the confident samples from unlabeled data for pseudo-labeling. Our strategy for selecting the confident samples is the following step(Figure 2). We regard the outputs of the trained model as the first arrival probability. At first, we select some fixed percent p1 of the predicted arrival times in descending order of arrival probability (from now on, we call this selected arrivals time "good picking time"). Next, we calculate the piece-wise linear regression curve of "good picking time" versus the offset between the source and receiver. Then, delete all predicted samples located more than three standard deviations away from the regression line. In addition to regression filtering, we remove some fixed percent p2 ($p1 < p2$) of the predicted traces in ascending order of first arrival probability. The remaining predicted first arrivals times are assigned the pseudo label. In addition to removing the outlier, we apply moveout correction for unlabeled data using piece-wise regression curve shift and limit the area of interest to 256 samples (Figure 3). This constraint helps increase the generalization ability of the trained model.

Multi-domain network learning

To further improve the preliminary first arrivals pick from the single-domain U-net based network, we use the multi-domain neural network based on 2D U-net architecture. The

supervised data are pseudo-labeled test dataset samples generated in the previous step. For multi-domain learning, we extract a series of adjacent seismic traces, which have all been assigned the pseudo label. Because the number of assigned pseudo-label samples is limited, we choose the number of adjacent traces small to ensure the supervised data volume. In this study, we extract five adjacent moveout-corrected and time-windowed traces as the supervised data. The input shape is 5 x 256. As Duan & Zhang (2020) also adopt the five adjacent traces for the input image to identify the poor pick, such a limited spatial input size may be enough to learn the horizontal consistency of seismic data. Because the optimized weight of convolutional kernels of the single-domain network helps extract the crucial feature of the first arrivals, we copy and expand weights from 1D into 2D and transfer them as the weight for the 2D U-net architecture. This weight transfer operation stabilizes multi-domain learning and reduces computational costs. We use Rectified Adam (RAdam) (Liu et al., 2019) as the optimization algorithm that schedules the learning rate warmup strategy (starts with a small learning rate and gradually increases the learning rate as the training progresses). During the workflow, stages 2) and 3) can be iterative; we can assign the new pseudo-label using the multi-domain learning prediction results. The self-trained algorithm works well if the assigned pseudo-label is correct and the unlabeled data contains the additional information on the actual first arrival times. However, incorrect pseudo-labels cause inappropriate training, accumulating errors in the iterative self-training process. Therefore, we should monitor the error between the predicted picking time and the “good picking time” (a confidently predicted time of single-domain model result, not a ground truth) at each iteration and stop the training if the error increases.

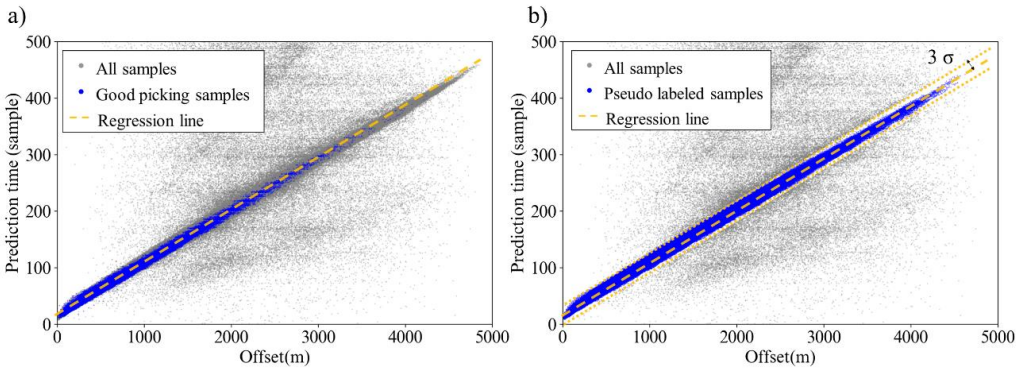

Figure 2. (a) Gray circles represent the preliminary predicted picks of the single-domain learning model, and blue circles are the confident picking (selected p1 % of whole samples) for calculating the pice-wise regression line(orange line). (b) The samples located more than three standard deviations away from the regression line (gray circle) are eliminated. Besides, the remains with the below arrivals probability of p2 percentile are removed. Finally, the remains (blue circle) are used as pseudo-labeled data.

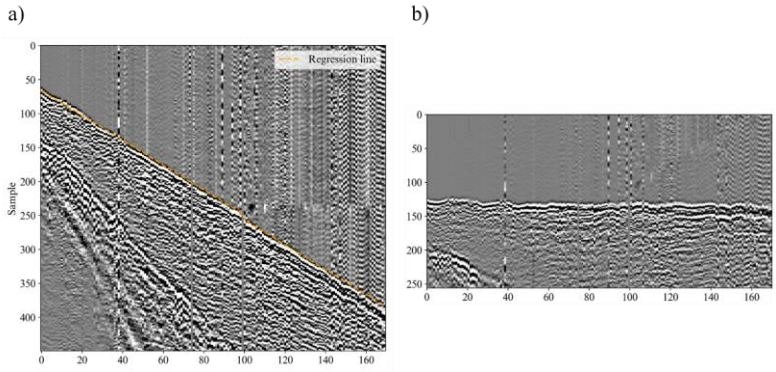

Figure 3. (a) A shot record before moveout correction. The dashed orange line indicates the piece-wise regression line. (b) A shot record after moveout correction and cropping.

EXPERIMENT & RESULT

St-Charles et al. (2021) recently released the open multi-survey dataset with manual picking annotation and U-Net based pickup accuracy benchmark. The dataset contains four distinct seismic explorations at hard rock mining sites: These are called “Lalor,” “Brunswick,” “Halfmile,” and “Sudbury.” Along with the multi-survey dataset, the first break pick annotations for 72.3 % of traces are provided. The annotations are deemed valid by a human expert. The accuracy benchmark, which we will compare our model results with, is based on a U-net architecture. The input to the benchmark model is a line gather which is treated as an image with multiple channels. Its first channel is seismic amplitude. Other channels contain source-receiver distances and the distances to the closest previous and next receivers on the same line.

For the performance metric, we use the Mean Average Error(MAE), the Mean Bias Error(MBE), and the Root Mean Squared Error (RMSE), where all errors are measured at the sample scale following St-Charles et al. (2021). We also calculate the Hit Rate at δ samples (HR@δpx) defined by the following equation,

$$\text{HR@}\delta\text{px} = \frac{n\{(|P_A - P_M| < \delta) \cap (P_A > 0)\}}{n(P_A > 0)} \quad (1)$$

where $P_M = \{P_{M_i}\}_{i=1}^{N}$ and $P_A = \{P_{A_i}\}_{i=1}^{N}$ are the manual and the automated pickings, respectively. N is the number of all traces. Additionally, we set the first arrivals times of no-pickup traces as -1 for both manual and automated picking. Hence, the equation (1) means the fraction of manually annotated traces where the prediction error is smaller than δ samples. In this paper, we set δ = 1, 3, 5, 7, 9 for the equation to evaluate the model capability under different accuracy requirements.

As per St-Charles et al. (2021), we conduct a random hyperparameter search of 50 trials in the single-domain network training phase. For each trial, we train on two sites and evaluate on a third site as the validation. Table 1 shows the cross-validation folds setting. During the hyperparameter search, we set the batch size for training at 300. Table 2 displays the explored range of the hyperparameters. For the optimization, we use learning rates uniformly across a logarithmic scale of $[10^{-5}, 5 \cdot 10^{-3}]$. Early stopping is performed if the HR@1px on the validation site does not improve for more than 5 consecutive epochs. [

Table 1. Distribution over hyperparamters for 50 trials of single-domain learning

Hyper-parameters	Range
Iterations of Hyperparameter Search	50
Initial Learning Rate	10^{-5}~5×10^{-3}
Number of Feature map at First CNN Layer	16, 32 ,64
Number of CNN Layer	8, 10
Dropout ratio	0~0.5

Table 2. The cross-validation setting

Fold No	Training Set	Validation Set	Test Set
1	Sudbury, Halfmile	Lalor	Brunswick
2	Halfmile, Lalor	Brunswick	Sudbury
3	Brunswick, Sudbury	Halfmile	Lalor
4	Lalor, Brunswick	Sudbury	Halfmile

The schematic U-net architecture of our method is illustrated in Figure 4. We set either 3 or 4 down/up-sampling stages in the model. Each stage consists of a flat convolution layer and a down/up-sampling convolution layer. The flat convolution layer has a kernel with a size of 3 and a convolution step size of 1. The down-sampling convolution layer has a kernel with a size of 7 and a convolution step size of 4. The up-sampling is done by deconvolution. At each up/down-sampling step, we double/halve the number of feature maps. The first convolution layer contains

either 16, 32, or 64 feature maps. After each convolution map, the convolution output is applied to batch normalization (Ioffe and Szegedy, 2015) and dropout(Srivastava et al., 2014), and then it passes a rectified linear unit (ReLU). Dropout ratios are set uniformly in the range [0~0.5].

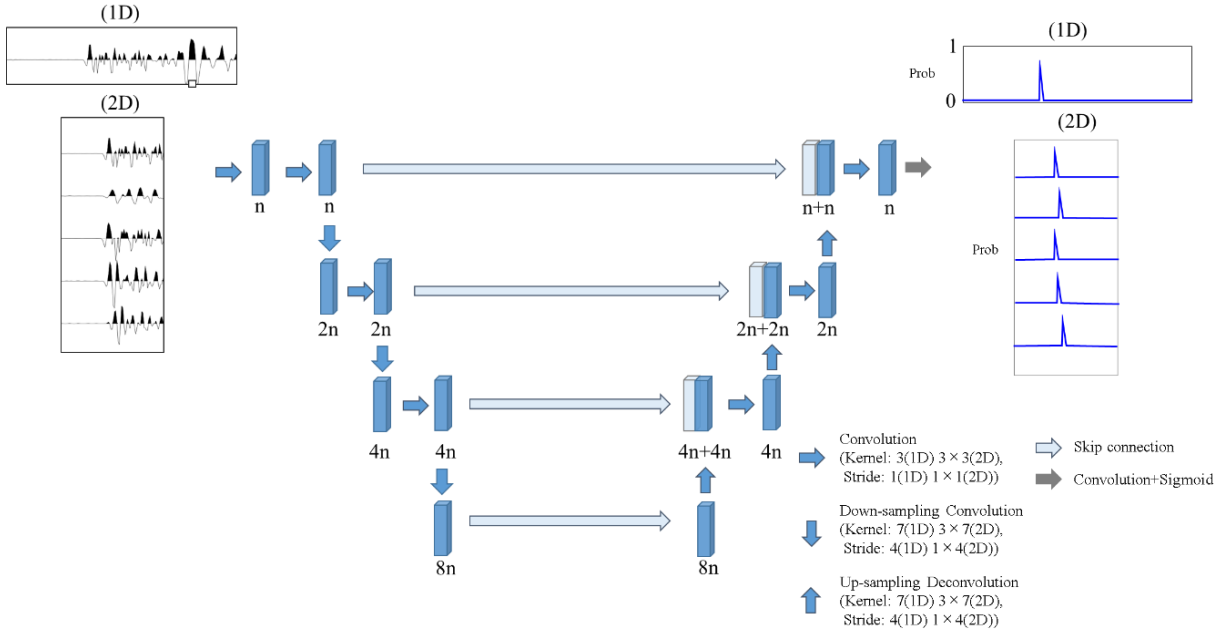

Figure 4. The schematic U-net architecture in this study. “n” stand for the number of feature maps of the first convolution layer. During the hyperparameter search, we search the number of 3 or 4 down/up-sampling stages in the model, and this figure shows the case of the 3 down/up-sampling stages. Except for the last convolution layer, the convolution outputs are applied to batch normalization and dropout, passing a rectified linear unit (ReLU). The output gives the arrivals probability for each sample in the input.

After 50 trials, we preserve the hyperparameter configuration leading to the model with the best validation HR@1px score. The 10 single-domain models have trained again with this configuration, starting from different random seeds. Then, we use each predicted result to assign the pseudo-label samples for 10 multi-domain training. We set the proportions of confident traces p1 to 10 %. As the ratio of eliminated traces p2, we use the 1 minus accuracy of the

validation score at the first iteration, and then the ratio p2 decreases 5 % by each iteration. The pseudo-labeled traces are applied flip augmentation with a spatial axis to increase the data volume and diversity. Multi-domain training uses the same hyperparameter configuration as single-domain training, except that the convolutional kernel dimension is converted from 1D to 2D. We extend the weights of the single-domain model to 2D and transfer them into the initial weights of the kernels. During the training phase, the batch size is fixed at 1000. We monitor the HR@1px score between the predicted picking time and the "good picking time" at each iteration of multi-domain learning and stop the training if the error does not decrease 2 consecutively iterations.

Figure 5 shows the distribution of our single-domain model (Light blue bar) and the benchmark (St-Charles et al., (2021)) (Blue dotted open bar) performance obtained from different validation sites during the hyperparameter search. We can observe that performance of our single-domain learning model is relatively stable at all test sites compared with the benchmark performance. Table 3 shows the evaluation metrics of the best trial for each fold computed on the test site. The single-domain learning model shows the intimate performance of the benchmark in terms of HR1px at test sites except for Lalor. It remarks better results on the Lalor test site, which has data with a different sampling rate which the predictive model did not experience during training and validation. The results indicate the high generalized ability of single-domain machine learning. However, the single-domain learning model performs slightly worse on the lenient accuracy requirement (HR@δpx, $\delta > 1$). This poor result may be due to the lack of spatial information of shot gather during the training. Multi-domain learning improves all of the model performance from single-domain results. As a result, our proposed method remarks the state-of-the performance except for some Hit Ratio of Brunswick site.

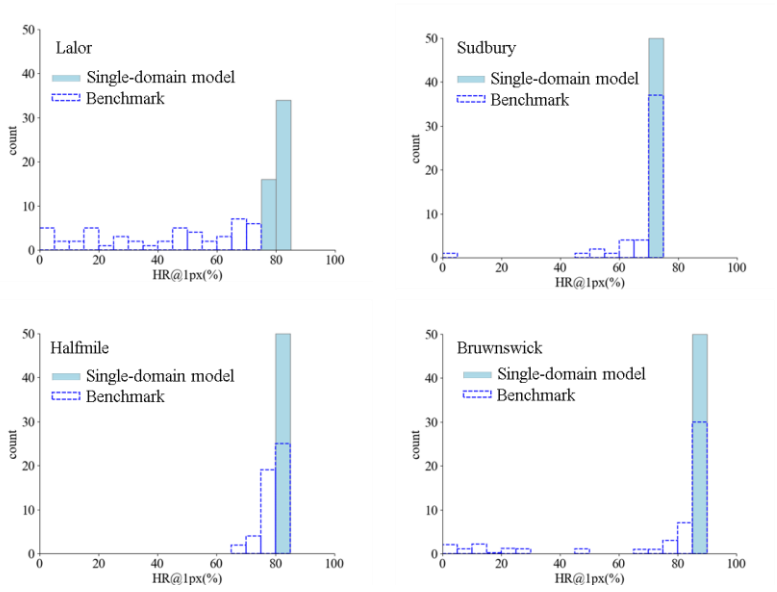

Figure 5. Distribution of HR@1px scores obtained on the validation site of each fold across. Our single-domain model (light-blue bar) remarks a more consistent result than the benchmark result (blue dotted open bar).

Table 3. Metrics calculated on the test site. Ten runs with different random seeds were performed, and the mean plus-or-minus the standard deviation over the ten runs was shown. The Hit Rates(HR) are in percentage, and the errors are in a number of samples.

Site	Model	HR@1px	HR@3px	HR@5px	HR@7px	HR@9px	RMSE	MAE	MBE
Brunswick	Benchmark	87.6±1.4	**96.4±0.6**	**97.8±0.6**	**98.3±0.6**	**98.6±0.6**	50.2±13.8	4.5±2.3	3.8±2.5
	Single Phase-net	88.3±0.3	93.5±0.2	94.5±0.2	94.9±0.2	95.1±0.2	41.0±0.9	7.2±0.3	-2.3±0.7
	Proposed Method	**89.1±0.2**	96.1±0.2	97.5±0.2	98.1±0.2	**98.6±0.1**	**4.0±0.5**	**0.5±0.0**	**0.2±0.0**
Sudbury	Benchmark	73.1±0.5	93.9±0.6	96.2±0.6	97.5±0.5	98.2±0.5	35.1±12.3	2.8±1.2	1.2±1.1
	Single Phase-net	72.5±0.2	92.3±0.3	94.0±0.4	94.9±0.4	95.4±0.4	35.3±2.8	5.2±0.5	4.9±0.5
	Proposed Method	**74.4±0.3**	**94.8±0.3**	**96.9±0.3**	**97.9±0.3**	**98.5±0.3**	**3.3±0.2**	**0.7±0.1**	**0.1±0.1**
Lalor	Benchmark	76.3±1.8	80.0±1.7	82.7±1.7	86.4±1.9	89.0±2.0	460.0±74.0	124.0±40.0	123.0±40.0
	Single Phase-net	83.7±0.5	85.8±0.5	88.8±0.4	93.0±0.3	95.6±0.3	14.3±2.6	1.8±0.2	0.8±0.1
	Proposed Method	**86.4±0.2**	**87.9±0.2**	**90.5±0.2**	**94.2±0.1**	**96.7±0.1**	**4.6±0.1**	**1.0±0.0**	**0.6±0.1**
Halfmile	Benchmark	**83.8±0.5**	**92.6±0.5**	95.9±0.6	97.9±0.6	98.8±0.6	35.2±33.3	3.8±4.3	2.9±4.1
	Single Phase-net	83.2±0.2	89.9±0.1	93.0±0.1	95.2±0.1	96.3±0.1	24.9±0.6	3.6±0.1	1.4±0.2
	Proposed Method	**83.8±0.5**	**92.6±0.1**	**96.0±0.1**	**98.2±0.1**	**99.2±0.0**	**2.3±0.4**	**0.5±0.0**	**0.0±0.0**

Multi-stage segmentation picking Network (MSSPN) (Wang et al., 2022) also shows outstanding results against the first arrival picking of the Hardrock seismic multi-survey dataset. Through their workflow, they calibrated the human-labeled first arrival time, which means they moved the label of non-peak (non-trough) times to the nearest peak (trough) time. Not only their sophisticated workflow, but this operation also contributes to improving their result. However, because the first arrivals phase is easily affected by noise and energy absorption (Hu et al., 2019), the actual first arrivals time is sometimes not located at the peak or trough. Therefore, we decided not to calibrate the human-labeled first arrival time on purpose for comparison with benchmark results. In addition, they eliminated outliers from the result of their deep learning model. The post-processing improved their precision score (The proportion of the predicted samples are identified correctly). Note if there are no outlier exclusions, precision is equal to HR@1px. Suppose we remove outliers according to the arrival probability, our precision score increases while our recall score (the proportion of the manually annotated samples are identified correctly) decreases. Figure 6 shows the precision-recall curve of our proposed method with picking calibration operations. Unfortunately, because we cannot find the recall from MSPNN results, we cannot compare our results with theirs.

DISCUSSION

Due to the enormous hyperparameter space and computation costs, we cannot guarantee that our hyperparameter choices (such as the number of traces of inputs of a multi-domain network, the proportion of selecting confident samples (p1, p2), and neural network architecture) are optimal. However, our result is enough to reveal the effectiveness of self-trained learning in which a multi-domain network learns from the predicted arrivals times of a single-domain network.

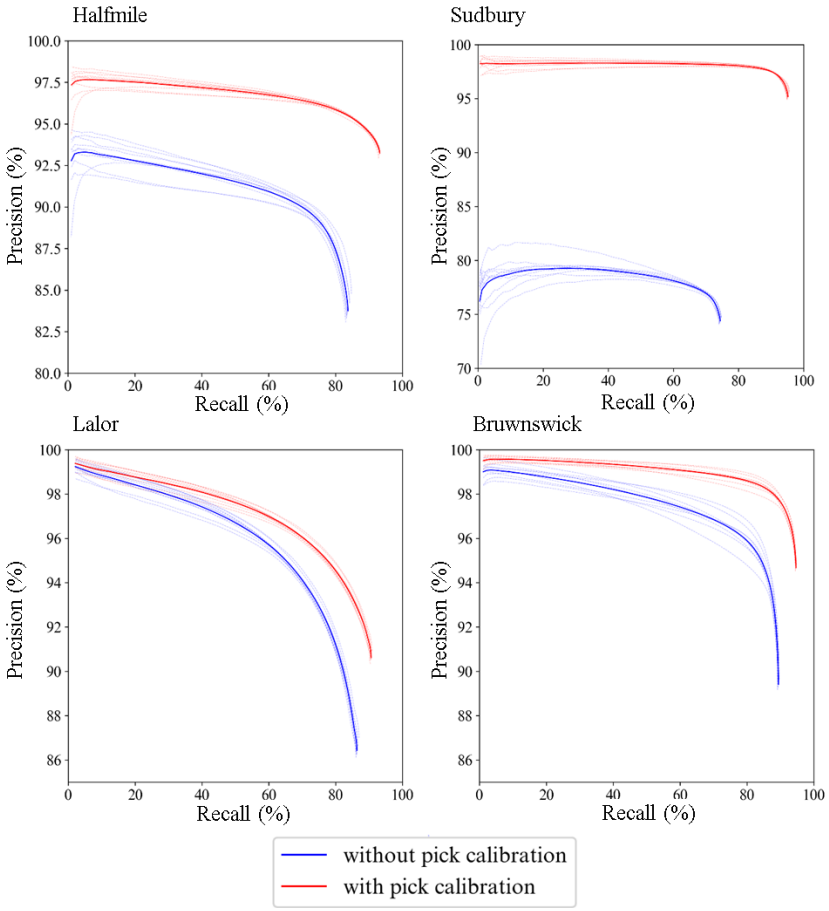

Figure 6. Precision-Recall curve of our proposed method. Results with pick calibration (Red lines) are better than the ones without pick calibration. The dashed shaded lines represent each random seed's results, and the solid lines represent their means.

We cannot assure whether analyzing seismic data as an image using a U-net based network is ideal. Mousavi et al. (2020) proposed a hierarchical attention-based network for simultaneous earthquake detection and phase picking. The attention mechanism helps to incorporate global and local scale features within the full waveform. It has the potential to improve picking accuracy and learning efficiency.

Because our proposed methods solely use the simple piece-wise linear regression for estimating the approximation relation between first arrivals time and offset, the proposed methods might have limitations for the more complex geological area having anisotropic and heterogeneous velocity. In such fields, common midpoint-offset-azimuth bins based on statistical analysis (Colombo et al., 2016) are an alternative to remove the outliers.

There may be other effective methods of eliminating outlier values during the self-training. For example, suppose we apply multi-domain prediction to other domains besides the common shot domains, such as the common receiver domain. In that case, we can consider the samples where the predicted first arrival times between multi-domains are consistent are highly confident. The ensemble method has the potential to increase predictive performance and model robustness. Furthermore, Han et al. (2021) applied the 3D U-Net for picking first arrivals and obtain a more consistent result than the one predicted by the 2-D. Although manually making the 3D labeled data is difficult and time-consuming, our method can generate 3D pseudo-labeled data using the trained single-or multi-domain network. Extending our approach to the 3D domain is the future work.

CONCLUSION

In this paper, we propose a single to multi-domain self-trained network for automatically picking the first arrivals in seismic data. First, our network starts learning from single-domain with labeled data to generate the pseudo-label on unlabeled data. Then our network learns on multi-domain using pseudo-label as supervised data. The hybrid domain learning strategy derives the advantage from both single- and multi-domain; the generalizability of the single-domain learning and noise robustness of the multi-level learning. In experiments on an open multi-survey dataset, our workflow clarifies the multi-analysis improves the preliminary picking result of

single-domain learning. Besides, the result of the Lalor site, which has a sampling rate that differs from the training data, indicates the high generalizability of our model. Finally, our proposed method surpasses the benchmark on most evaluation metrics.

ACKNOWLEDGEMENT

The author would like to thank Mr Pierre-Luc St-Charles from Applied Machine Learning Research Team Mila, Qu'ebec AI Institute for providing the multi-survey open datasets.

REFERENCE

Akram, J., & Eaton, D. W. (2016). A review and appraisal of arrival-time picking methods for downhole microseismic dataArrival-time picking methods: Geophysics, **81(2)**, KS71–KS91.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. Bulletin of the Seismological Society of America: **68(5)**, 1521–1532. https://doi.org/10.1785/BSSA0680051521

Boschetti, F., Dentith, M. D., & List, R. D. (1996). A fractal-based algorithm for detecting first arrivals on seismic traces. Geophysics: **61(4)**, 1095–1102.

Chen, Y., Zhang, G., Bai, M., Zu, S., Guan, Z., & Zhang, M. (2019). Automatic Waveform Classification and Arrival Picking Based on Convolutional Neural Network. Earth and Space Science: **6(7)**, 1244–1261. https://doi.org/10.1029/2018EA000466

Colombo, D., Miorelli, F., Sandoval, E., & Erickson, K. (2016). Fully automated near-surface analysis by surface-consistent refraction method: Geophysics, **81(4)**, U39–U49.

Coppens, F. (1985). First arrival picking on common-offset trace collections for automatic estimation of static corrections: Geophysical Prospecting, **33(8)**, 1212–1231.

Diehl, T., Deichmann, N., Kissling, E., & Husen, S. (2009). Automatic S-wave picker for local earthquake tomography: Bulletin of the Seismological Society of America, **99(3)**, 1906–1920.

Duan, X., & Zhang, J. (2020). Multitrace first-break picking using an integrated seismic and machine learning methodPicking based on machine learning: Geophysics, **85(4)**, WA269–WA277.

Fralick, S. (1967). Learning to recognize patterns without a teacher: IEEE Transactions on Information Theory, **13(1)**, 57–64.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 580–587.

Han, L., Wong, J., & Bancroft, J. C. (2009). Time picking and random noise reduction on microseismic data: CREWES Research Report, **21**, 1–13.

Han, S., Liu, Y., Li, Y., & Luo, Y. (2021). First Arrival Traveltime Picking Through 3-D U-Net: IEEE Geoscience and Remote Sensing Letters, **19**, 1–5.

Hu, L., Zheng, X., Duan, Y., Yan, X., Hu, Y., & Zhang, X. (2019). First-arrival picking with a U-net convolutional network: Geophysics, **84(6)**, U45–U57.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift: International Conference on Machine Learning, 448–456.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization: ArXiv Preprint ArXiv:1412.6980.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks: Communications of the ACM, **60(6)**, 84–90.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition: Neural Computation, **1(4)**, 541–551.

Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J., & Han, J. (2019). On the variance of the adaptive learning rate and beyond: ArXiv Preprint ArXiv:1908.03265.

Loginov, G. N., Duchkov, A. A., Litvichenko, D. A., & Alyamkin, S. A. (2022). Convolution neural network application for first-break picking for land seismic data: Geophysical Prospecting, **70(7)**, 1093–1115.

McCormack, M. D., Zaucha, D. E., & Dushek, D. W. (1993). First-break refraction event picking and seismic data trace editing using neural networks: Geophysics, **58(1)**, 67–78.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking: Nature Communications, **11(1)**, Article 1. https://doi.org/10.1038/s41467-020-17591-w

Murat, M. E., & Rudman, A. J. (1992). Automated first arrival picking: A neural network approach1: Geophysical Prospecting, **40(6)**, 587–604.

Ozawa, M., Kawasaki, S., Higashinaka, M., Sato, H., & Abe, S. (2018, March 14). Detecting seismic first arrivals using deep Convolution Neural Networks: Japan Geoscience Union. https://confit.atlas.jp/guide/event/jpgu2018/subject/STT50-01/detail

Palmer, D. (1980). The generalized reciprocal method of seismic refraction interpretation: Society of Exploration Geophysicists.

Peraldi, R., & Clement, A. (1972). Digital Processing of Refraction Data Study of First Arrivals*. Geophysical Prospecting: **20(3)**, 529–548. https://doi.org/10.1111/j.1365-2478.1972.tb00653.x

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation: International Conference on Medical Image Computing and Computer-Assisted Intervention, 234–241.

Sabbione, J. I., & Velis, D. (2010). Automatic first-breaks picking: New strategies and algorithms: Geophysics, 75(4), V67–V76.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting: The Journal of Machine Learning Research, 15(1), 1929–1958.

St-Charles, P.-L., Rousseau, B., Ghosn, J., Nantel, J.-P., Mila, Q. A., & Schetselaar, G. B. E. (2021). A Multi-Survey Dataset and Benchmark for First Break Picking in Hard Rock Seismic

Exploration: in Proc. Neurips 2021 Workshop on Machine Learning for the Physical Sciences (ML4PS).

Stevenson, P. R. (1976). Microearthquakes at Flathead Lake, Montana: A study using automatic earthquake processing: Bulletin of the Seismological Society of America, **66(1)**, 61–80.

Takanami, T., & Kitagawa, G. (1991). Estimation of the arrival times of seismic waves by multivariate time series model: Annals of the Institute of Statistical Mathematics, **43(3)**, 407–433. https://doi.org/10.1007/BF00053364

Tsai, K. C., Hu, W., Wu, X., Chen, J., & Han, Z. (2018). First-break automatic picking with deep semisupervised learning neural network: In SEG Technical Program Expanded Abstracts 2018, 2181–2185. Society of Exploration Geophysicists.

Wang, H., Zhang, J., Wei, X., Zhang, C., Guo, Z., Long, L., & Wang, Y. (2022). MSSPN: Automatic First Arrival Picking using Multi-Stage Segmentation Picking Network: ArXiv Preprint ArXiv:2209.03132.

Yarowsky, D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. 33rd Annual Meeting of the Association for Computational Linguistics, 189–196.

Yuan, P., Wang, S., Hu, W., Wu, X., Chen, J., & Van Nguyen, H. (2020). A robust first-arrival picking workflow using convolutional and recurrent neural networks: Geophysics, **85(5)**, U109–U119.

Yuan, S., Liu, J., Wang, S., Wang, T., & Shi, P. (2018). Seismic waveform classification and first-break picking using convolution neural networks: IEEE Geoscience and Remote Sensing Letters, **15(2)**, 272–276.

Yung, S. K., & Ikelle, L. T. (1997). An example of seismic time picking by third-order bicoherence: Geophysics, **62(6)**, 1947–1952.

Zhu, W., & Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method: Geophysical Journal International, **216(1)**, 261–273. https://doi.org/10.1093/gji/ggy423