Research on seismic wave first arrival Picking Based on improved U-Net

Ping Liu¹*, Mengrou Yao¹, Zhenhua Han¹, Yanrong Li², Xinrui Li¹ and Juncai Shi¹

¹ College of Data Science, Taiyuan University of Technology, Jinzhong, Shanxi Province, 030600, China
² College of Mining Engineering, Taiyuan University of Technology, Jinzhong, Shanxi Province, 030600, China
*Corresponding author’s e-mail: liuping01@tyut.edu.cn

Abstract: For the current problem of low accuracy of seismic wave phase pickup, this paper proposes a method based on the improved U-Net seismic wave P-wave and S-wave first-to-arrival pickup. Firstly, in order to reduce the impact of pre-processing operations such as filtering, the 20,000 three-component seismic waveform data will be normalized as input. Secondly, in order to reduce the loss of information during the seismic waveform extraction and increase the amount of information of effective features at the starting point, the Residual Unit is added to the encoder part of U-Net only, instead of adding it to the decoder part. In order to reduce the loss of high frequency information of seismic waveform, convolution operation is used instead of pooling operation. Finally, the probability distributions of P-wave, S-wave and noise are used as the output. The results show that the accuracy of the first-to-arrival pickup of P-wave is 93% and the accuracy of the first-to-arrival pickup of S-wave is 90% of the test set. The improved U-Net increased 30.9% and 6.2% in P-wave accuracy, 7.6% and 2.8% in S-wave accuracy, 39.6% and 4.0% in the P-wave recall, and 41.5% and 4.7% in the S-wave recall, respectively, compared with the conventional STA/LTA and U-Net. The improved U-Net outperforms the other two models in both evaluation metrics of recall and accuracy. The research results of this paper are of great practical significance in the localization of earthquakes, the interpretation of the earthquake gestation mechanism, and the prevention and mitigation of earthquakes.

1. Introduction

Seismic monitoring and positioning is the foundation of seismology, and the quality of the seismic catalog depends largely on the accuracy of seismic wave arrival time measurements, which are usually derived from judgments made by seismologists based on years of work experience. With the intensive deployment of seismic networks, the speed and scale of geophone deployment has been increasing, and the amount of seismic waveform data has increased dramatically. The traditional method of picking up P-wave and S-wave first arrivals require a lot of manpower and time, while the automated seismic phase identification and first arrival pickup can efficiently process the seismic data. Accurate identification of P- and S-wave first arrivals is of great practical importance for earthquake localization, seismogenesis interpretation, and disaster prevention and mitigation.

Early seismic wave P- and S-wave first-to-first pickups were performed entirely manually, and with the development of technology, automated seismic phase pickup techniques have emerged. The
short-term average/long-term average (STA/LTA) [1,2] method used the ratio of the average of the short-time window to the long-time window of the seismic wave signal to describe the energy variation of this segment of the signal. The STA/LTA method is simple in principle, fast in calculation, and can quickly detect obvious seismic events, but it is greatly influenced by the signal-to-noise ratio, threshold value and the selection of feature function. Baer & Kradolfer (1987) [3] improved the STA/LTA method using the envelope as the characteristic function. Sleeman & van Eck (1999) [4] applied joint autoregressive (AR) modelling of the noise and seismic signal and used the Akaike Information Criterion to determine the onset of a seismic signal. Gentili & Michelini (2006) [5] tested a traditional shallow neural network to select the first-to-arrival times of P and S waves based on four manually defined features: variance, absolute value of skewness, kurtosis, and combined skewness and kurtosis predictions based on a sliding window. The traditional automated method used manually defined features (e.g., time, amplitude, polarization angle, etc.), which required prior extraction of the features of the seismic waveforms, followed by data pre-processing (e.g., band-pass filtering and setting activation thresholds). Different processing means extract different features and obtain different results, which makes the compilation of the seismic catalog extremely complicated.

With the development of science and technology, deep learning has made great progress, and new network structures such as convolutional neural networks [6-13] have been introduced, and their use in earthquake seismic phase pickup has created a boom. Perol T, et al. (2018) [14] applied convolutional neural networks to the study of seismic phase recognition and designed an 8-layer convolutional neural network—ConvNetQuake, to recognize continuous waveforms of earthquakes in the Oklahoma region of the United States, and proved that this network has good recognition effects. But the amount of data used by ConvNetQuake also reached 700,000 data and was not friendly for low-level seismic wave recognition. Zhao et al. (2019) [15] used a neural network to identify seismic phases, and the results were better than those of STA/LTA on continuous waveforms. Li Jian et al. (2020) [16] used a migration learning approach to train 2 million seismic data from the US Southern California seismic network to provide a new idea for earthquake phase recognition.

Based on the above analysis, the traditional automatic pickup methods have poor generalization, and different parameter and threshold choices have a great impact on the results. For seismic signals with low signal-to-noise ratio, the detection effect often fails to achieve the expected results. With the rise of deep learning, the method of CNN has been more applied in the past two years, and CNN has been better applied in the problems of seismic and noise classification. However, its accuracy needs to be further improved in the first-to-arrival pickup of seismic waves. In this paper, we present a method based on an improved U-Net model for first-to-arrival pickup of P-waves and S-waves.

2. Data

In this paper, we used seismic waveform data from the Institute of Geophysics of the China Seismological Bureau for the China track of the 2020 Microsoft Innovation Cup in October 2019 for the preliminary round of the seismic project. The dataset contains about 20,000 seismic events that have been detected and confirmed by seismologists. The various types of data in the dataset are shown in Figure 1. The length of seismic waveform data in the dataset is 30 seconds, the sampling rate is 100 HZ, and the size of each data segment is 3000 sampling points. The organization of the individual data is (3000, 3), where 3000 represents the total number of sampling points and 3 represents the number of channels (three channels are vertical, east-west and north-south directions). The three-component seismic waveform data in the dataset are shown in Figure 2. The initial arrival times of the P and S waves of the seismic waveform data are marked by seismologists. According to the catalog of earthquake phases provided by seismologists, 25 sampling points before and after the P-wave arrival time are taken as the range of P-wave arrival time, and 50 sampling points before and after the S-wave arrival time are taken as the range of S-wave arrival time. The samples involved in the training and their labels are shown in Figure 3.
Figure 1. Various types of data in the Dataset

Figure 2. Three-component waveform data in the Dataset

Figure 3. Samples participating in training and their labels
(i) represents the label of the noise; (ii) represents the label when the P wave arrives; (iii) represents the label when the S wave arrives; (iv) represents the original seismic waveform data in a certain direction. The red vertical line represents the first arrival of the marked P wave, and the blue vertical line represents the first arrival of the marked S wave.

3. Method

3.1. U-Net model
The U-Net model [17] mainly borrows the idea of full convolutional neural network (FCN)[18] and adopts the structure of "encoder-decoder". The encoder is used to obtain the contextual information and the decoder is used to pinpoint the location, and the two structures are symmetric to each other. Firstly, the U-Net model uses Conv (+BN+ReLU)+POOLing to downsample the image several times. Then, the image is upsampled and the low-level feature image before cropping is cropped. It is fused with the upsampled feature image, and the process of upsampling and fusion is repeated until a segmentation image with the same size as the input image is obtained. The U-Net network structure is shown in Figure 4. The final feature map is obtained by combining with the cross-entropy loss.
function, and the Softmax function is used to calculate the final probability distribution. The Softmax function is defined as follows:

\[ p_k(x) = \exp(a_k(x)) \left( \sum_{k=1}^{K} \exp \left( a_k'(x) \right) \right) \]

where \( a_k(x) \) represents the activation function of the \( k \)th feature map at the \( x \)th pixel point. The \( K \) represents the number of categories and \( p_k(x) \) represents the maximum function. The cross-entropy function is defined as follows:

\[ E = \sum_{x \in \Omega} \omega(x) \log \left( p_{l(x)}(x) \right) \]

where \( l = \{1, \ldots, K\} \) represents the correct label for each pixel.

3.2. Residual Unit model

Residual Unit, or residual learning unit, was proposed by Kaiming He [19] in 2015. ResNet model is composed of a series of residual units superimposed. The structure of Residual block residual learning unit is shown in Figure 5. Residual learning unit is defined as follows:

\[ L_i = h_i(x_i) + F(x_i + W_i) \]

\[ x_{i+1} = f(L_i) \]

where \( x_i \) and \( x_{i+1} \) represent the input and output layers of the current residual unit. The \( F(x) \) function represents the residual mapping function, and \( W \) represents the weight matrix. This design is because constructing the mapping function \( h(x) \) is equivalent to constructing the residual mapping function \( F(x) \), and the residual mapping function is easier to optimize.

3.3. The improved U-Net model

When using U-Net model for seismic P-wave and S-wave initial arrival pickup, the most important thing is to find the starting points of P-wave initial arrival and S-wave initial arrival, so the features of the starting point of P-wave and S-wave initial arrival need to be extracted. The accuracy and recall of traditional U-Net for seismic P-wave and S-wave initial arrival pickup are relatively low. And feature information is needed more in the seismic initial arrival pickup task than in the simple seismic waveform detection task, which is expressed as the difference between time and seismic waveform amplitude in the time domain. Although the traditional "encode-decode" structure can simultaneously identify seismic phases and pick them up at first arrival, feature information is still lost in the process of downsampling. The pooling layer used in the downsampling process loses the high frequency features of the seismic information. To address this problem, the pooling layer is replaced by a
convolutional layer to reduce the loss of high-frequency seismic information caused by the pooling operation. The improved U-Net model allows simultaneous seismic phase identification and localization, replicating seismic signal features from the lower layers to the corresponding higher layer features. This creates new paths for the propagation of seismic waveform features and makes it easier to propagate seismic signals between lower and higher levels, allowing the fine features at lower levels to complement the semantic features at higher levels. The residual learning simplifies the training of the neural network. And the residual units and the skip-connection in U-Net allow the low-level information of the seismic waveforms to be well combined with the abstract high-level information so that the information does not degrade during propagation. The improved U-Net model uses a 9-layer network architecture, and the structure of the model is shown in Figure 6.

![Figure 6. The improved U-Net network structure diagram](image)

The model includes three parts: encode, bridge, and decode. The encode part is to acquire the feature information of the image. The decode part is to recover the feature information to the original data size, and the encode and decode parts are connected by the bridge part. Both encode and bridge parts are composed of a convolutional block with a convolutional kernel size of 7 and a residual learning unit. Each convolutional block consists of a convolutional layer, a BN layer, a ReLU activation layer, and a Dropout layer, representing the input and output of the mapping connection unit. The encode part has four residual learning units, and in each learning unit, instead of using a pooling layer, a convolutional layer with a step size of 4 is used to reduce the feature mapping (in general, when using convolutional neural networks for classification problems, the pooling operation is used to extract low-level feature information, and since the key to picking up seismic wave P-wave and S-wave initial-to-arrival pickups is to find the data points where the seismic data undergoes abrupt changes, using the pooling operation will degrade the pickup accuracy, so the pooling operation is not used in this model, and the convolutional operation with a step size of 4 is used directly instead). Accordingly, the decode part also includes 4 learning units, and before each learning unit, the upsampling of the low-level feature mapping is connected to the feature mapping of the corresponding the encode part. After the last layer of the decode part, the multichannel feature mappings are projected into the desired segmentation using 1×1 convolution and Softmax layers. The network ends with a probability distribution output for P-wave, S-wave and noise using the Softmax function.

\[
p_i(x) = \frac{e^{x_i(x)}}{\sum_{k=1}^{K} e^{x_k(x)}}
\]

where \(i = 1, 2, 3\), represents the P-wave, S-wave and noise; \(z(x)\) represents the output of the last layer.

The loss function is usually a measure used to evaluate the degree of difference between the predicted and true values of a model. The smaller the value of the loss function, the smaller the difference between the predicted and true values of the model, and the better the model. In this paper, we use the
cross-entropy loss function to describe the difference between the predicted and true values of the seismic model. The loss function is defined as follows:

$$H(p, q) = - \sum_{i=1}^{3} \sum_{x} p(x) \log q_i(x)$$  \hspace{1cm} (6)

where $p(x)$ represents the true distribution of the seismic waveform data and $q(x)$ represents the predicted distribution of the seismic waveform data.

4. Experiments

4.1. Training and testing of the model

In this paper, the model was implemented based on Keras, and the optimization algorithms were stochastic gradient descent (SGD) algorithm and Adma algorithm [20]. And the learning rate of the model was set to $1 \times 10^{-3}$. The trained model was applied to the test set and the test results performed well. The results of training and testing were shown in Figure 7. When training was performed on the training set, the accuracy reached 97% at 20,000 steps and was maintained at this level for subsequent steps. On the test set, the recognition rate of P-wave increased from 85% to 93% and the recognition accuracy of S-wave increased from 81% to 90% when the step was increased from 20,000 to 40,000. In addition, the loss decreases from 10.34 to 0.33 from the beginning of training to 40,000 steps.

(a) The training loss result graph

(b) The picking accuracy graph at the first arrival of the P wave in the Training-set

(c) The picking accuracy graph at the first arrival of the S wave in the Training-set
The picking accuracy map at the first arrival of the P wave in the Test-set

The picking accuracy map at the first arrival of the S wave in the Test-set

Figure 7. Training and test results

4.2. Comparative model analysis

In this paper, the first-to-arrival times of seismic waves P-wave and S-wave are picked up using the improved U-Net model. It is also compared with the conventional STA/LTA and U-Net. The STA/LTA method is based on the principle of signal processing to pick up the first arrival times of seismic P-wave and S-waves by signal equation calculation. The U-Net uses the principle of image segmentation to determine the starting points of P-wave and S-waves. The experimental results are evaluated using the confusion matrix, which is defined as shown in Table 1. The comparison results of STA/LTA, U-Net and the modified U-Net for P-wave and S-wave first arrival time accuracy and recall are shown in Table 2.

### Table 1. Definition of confusion matrix.

| True Value | N   | P   | S   |
|------------|-----|-----|-----|
| Predicted value | N   | P   | S   |
| N          | $TP_{NN}$ | $FP_{NP}$ | $FP_{NS}$ |
| P          | $FP_{PN}$ | $TP_{PP}$ | $FP_{PS}$ |
| S          | $FP_{SN}$ | $FP_{SP}$ | $TP_{SS}$ |

The recall $R_P$ and precision $P_P$ at the first arrival of the P-wave are calculated as follows:

$$R_P = \frac{TP_{PP}}{TP_{PP} + FP_{PN} + FP_{NP}}$$  \hspace{1cm} (7)  

$$P_P = \frac{TP_{PP}}{TP_{PP} + FP_{PN} + FP_{NP}}$$  \hspace{1cm} (8)  

The recall $R_S$ and precision $P_S$ at the first arrival of the S-wave are calculated as follows:

$$R_S = \frac{TP_{SS}}{TP_{SS} + FP_{NS} + FP_{PS}}$$  \hspace{1cm} (9)  

$$P_S = \frac{TP_{SS}}{TP_{SS} + FP_{NS} + FP_{PS}}$$  \hspace{1cm} (10)  

Table 2. Comparison of STA/LTA, U-Net and the improved U-Net on the first arrival recall and precision results of P wave and S wave

| model         | Recall | Precision |
|---------------|--------|-----------|
|               | P   | S   | P   | S   |
| STA/LTA       | 45.4% | 39.5% | 58.3% | 76.9% |  
| U-Net         | 81%  | 76.3% | 83%  | 81.7% |  
| The improved U-Net | 85%  | 81%  | 89.2% | 84.5% |

As can be seen from Table 2, the two convolutional neural network methods, the U-Net model and the improved U-Net model, are higher than the traditional method, STA/LTA, in the two metrics of
recall and precision, indicating that the convolutional neural network method has better generalization and stronger model robustness. From the metric of precision, it can be concluded that for the precision of P-wave, the improved U-NeNet is 30.9% higher than STA/LTA and 6.2% higher than U-Net; for the precision of S-wave, the improved U-NeNet is 7.6% higher than STA/LTA and 2.8% higher than U-Net. From the metric of recall, it can be concluded that for the recall of P-wave, the improved U-NeNet is 39.6% higher than STA/LTA and 4% higher than U-NeNet, and for the recall of S-wave, the improved U-NeNet is 41.5% higher than STA/LTA and 4.7% higher than U-Net. Figure 8 shows some examples of the success of improved U-NeNet pickup of P- and S-wave initial arrivals.

4.3. Analysis of seismic event waveform applications in Shanxi Province
Model validation was performed for the event waveform of March 18, 2013 in Shanxi Province, for which the sampling rate was 100 HZ and the window length was set to 30 seconds, and the test results were compared with STA/LTA and U-NeNet. The validation results are shown in Figure 9.

Figure 9. Comparison of the picking effect of the same event waveform with the typical STA/LTA method. U-NeNet and the improved U-NeNet on the first arrival time of P-wave and S-wave selected by seismologists. In the figure, the red dotted line represents the first arrival of the P wave identified by the expert, and the blue dotted line represents the first arrival of the S wave identified by the expert.

From the analysis in Figure 11, it can be concluded that the error of STA/LTA for the first arrival pickup of P-wave is 1.25 s, and the error of U-NeNet model for the first arrival pickup of P-wave is 0.63 s. The error of the U-NeNet model for P-wave first arrival pickup is 0.63 s, and the error of the improved U-NeNet model for P-wave first arrival pickup is 0.63 s. The improved U-NeNet model has a pickup error
of 0.34 s for the P-wave initial arrival time. The error of STA/LTA for S-wave first-arrival pickup is 2.57s. The pickup error of the U-Net model for the first arrival time of S-wave is 0.56s. The improved U-Net model has a pickup error of 0.41s for S-wave initial arrival time, and the improved U-Net outperforms the typical STA/LTA and U-Net for both P-wave and S-wave first-arrival pickups. Compared with the traditional method STA/LTA with U-Net, the improved model performs well both in terms of the error of P-wave first-to-arrival pickup and the error of S-wave first-to-arrival pickup. This is because STA/LTA is strongly influenced by the threshold and signal-to-noise ratio, if the signal-to-noise ratio is too low it will lead to poor pickup, and a high threshold setting will also lead to poor pickup. The convolutional neural network, on the other hand, learns the waveform characteristics of the seismic data to identify and pick up, and is not particularly sensitive to the threshold and signal-to-noise ratio. Therefore, the convolutional neural network approach is more advantageous in a comprehensive view. However, when comparing P-wave first arrival with S-wave first arrival, S-wave is more difficult to pick up due to the influence of P-wave tail and mantle reflection, which leads to a larger pickup error than P-wave first arrival.

5. Conclusion
The experimental results in this paper show the promising application of convolutional neural networks for P-wave and S-wave first arrival pickup of seismic waves. The traditional STA/LTA method is very sensitive to the threshold selection of P-wave first arrival time and S-wave first arrival time, and requires a trade-off between too high and too low thresholds, which can easily lead to arrival time delay, while the improved U-Net model is relatively insensitive to the selection of thresholds. The traditional STA/LTA method must perform data pre-processing operations such as filtering, otherwise it will greatly reduce the pickup performance of STA/LTA, while the improved U-Net model does not perform data filtering operations, but only normalizes the data without other data pre-processing, which largely preserves the originality and integrity of the data, and can still pick up the P-wave and S-wave very well first-to-arrival time.

In this paper, we established a model based on the improved U-Net for seismic wave P-wave and S-wave first-to-arrival pickup. Firstly, the three-component seismic waveform data are only normalized as input to reduce the influence from pre-processing operations such as filtering. Secondly, the Residual Unit is added to the encoder part of U-Net only, and the low-level features are effectively combined with the high-level features to reduce the loss of information during the seismic waveform extraction. Finally, the pooling operation is used to replace the convolution operation to reduce the loss of seismic high-frequency information from the pooling operation. The improved U-Net model achieves an accuracy of up to 97% on the training set and 93% on the test set for the initial arrival time of P-wave and 90% for the initial arrival time of S-wave. Meanwhile, compared with the typical STA/LTA and U-Net, the improved U-Net model is higher than the other two models in both accuracy and recall rate. Finally, the waveform data of Shanxi Province on March 8, 2013 was selected to compare this paper's model with the typical STA/LTA. The pickup error of the improved U-Net model is significantly smaller than that of the typical STA/LTA. The improved U-Net model has shown superiority over the conventional method in the first-arrival time pickup of P- and S-waves. In the process of calculating the seismic phase arrival time, there is no need to set the relevant thresholds according to specific conditions, and the residual learning unit and skip connection allow the low-level information of seismic waveforms to be well combined with the abstract high-level information, so that the information will not be degraded during the propagation process. Accurate P- and S-wave initial-to-arrival pickups help to extract as much information as possible from the waveforms and help to detect earthquakes, which is expected to provide important practical implications for earthquake localization, interpretation of earthquake inception mechanisms, and earthquake prevention and mitigation.
References

[1] Allen R V. (1978). Automatic earthquake recognition and timing from single traces[J]. Bulletin of the Seismological Society of America, 68(5):1521-1532.

[2] Allen R V. (1985). Automatic earthquake recognition and timing from single traces[J]. Bulletin of the Seismological Society of America, 75(6):1827-1835.

[3] Li L, Zhang J.G, Li S.L. (2018). Developments of Research on Automatic Aftershock Recognition Technology. Journal of Seismological Research, 41(01):1-13+157. (in Chinese)

[4] Shelly D. R. (2017). A 15 year catalog of more than 1 million low-frequency earthquakes: Tracking tremor and slip along the deep San Andreas Fault. Journal of Geophysical Research: Solid Earth,122(5):3739–3753.

[5] Li Y, Gao Y, Ye B, Li X.B. (2017). Study on Auto Detection Method of Active Source Event Using Template Matching Technique[J]. Journal of Geodesy and Geodynamics, 37(07):752-755. (in Chinese)

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

[7] Szegedy C, Liu W, Jia Y, et al. (2015). Going deeper with convolutions[C]. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).Boston, MA, USA.1-9.

[8] Simonyan K, Zisserman A. (2014). Very deep convolutional networks for large-scale image recognition[J]. Computer Science. arXiv preprint arXiv:1409.1556.

[9] He K M, Zhang X Y, Ren S Q, et al. (2016). Deep residual learning for image recognition[J]. IEEE conference on computer vision and pattern recognition (CVPR). pp:770-778.

[10] Pourgharibshahi M, Lambert P. (2018). ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation[J]. arXiv, 67(8): 857-866.

[11] Szegedy C, Ioffe S, Vanhoucke V, et al. (2016). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[C]. The Thirty-First AAAI Conference on Artificial Intelligence (AAAI 2017). San Francisco, California, USA. 4: 12.

[12] Huang G, Liu Z, Pleiss G, Van Der Maaten L, Weinberger K. (2019). Convolutional Networks with Dense Connectivity [J]. IEEE Transactions On Pattern Analysis And Machine Intelligence, 1939-3539. doi: 10.1109/TPAMI.2019.2918284.

[13] Hu J, Li S, Samuel A, Gang S, Enhua W. (2018). Squeeze-and-excitation networks[J]. IEEE transactions on pattern analysis and machine intelligence, 42:2011-2023.

[14] Perol T, Gharbi M, Denolle M. (2018). Convolutional neural network for earthquake detection and location[J]. Science Advances, 4(2): e1700578, doi: 10.1126/sciadv.1700578.

[15] Zhao M, Chen S, Dave Y. (2019). Waveform classification and seismic recognition by convolution neural network[J]. Chinese Journal of Geophysics. 62(01): 374-382. (in Chinese)

[16] Li J, Wang X.M, Zhang Y.H, Wang W.D, Shang J, Gai L. (2020). Research on the seismic phase picking method based on the deep convolution neural network[J]. Chinese Journal of Geophysics. 63(04):1591-1606. (in Chinese)

[17] Ronneberger O, Fischer P, Brox T. (2015). U-net: Convolutional networks for biomedical image segmentation[J]. 9351:234-241. doi:10.1007/978-3-319-24574-4_28

[18] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2015.7298965

[19] He, K.M, Zhang, X.Y, Ren, S.Q, Sun, J. (2016). Deep Residual Learning for Image Recognition[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 770-778. doi:10.1109/cvpr.2016.90

[20] Kingma D., Ba J. (2015). Adam: A Method for Stochastic Optimization[C]. In: the 3rd International Conference for Learning Representations. San Diego.