Automatic Arrival Time Detection for Earthquakes Based on Stacked Denoising Autoencoder

Article *in* IEEE Geoscience and Remote Sensing Letters · August 2018

DOI: 10.1109/LGRS.2018.2861218

CITATIONS
17

READS
486

5 authors, including:

Omar M. Saad
National Research Institute of Astronomy and Geophysics
39 PUBLICATIONS **164** CITATIONS
SEE PROFILE

Ahmed Shalaby
Benha University
55 PUBLICATIONS **269** CITATIONS
SEE PROFILE

Lotfy Samy
National Research Institute of Astronomy and Geophysics
12 PUBLICATIONS **64** CITATIONS
SEE PROFILE

Mohammed Sharaf Sayed
Egypt-Japan University of Science and Technology
91 PUBLICATIONS **529** CITATIONS
SEE PROFILE

Automatic Arrival Time Detection for Earthquakes Based on Stacked Denoising Autoencoder

Omar M. Saad, K Inoue, Ahmed Shalaby, Lotfy Samy, and Mohammed S. Sayed.

***Abstract*—The accurate detection of P-wave arrival time is imperative for determining the hypocenter location of an earthquake. However, precise detection of onset time becomes more difficult when the signal-to-noise ratio (SNR) of the seismic data is low, such as during micro-earthquakes. In this paper, a stacked denoising autoencoder (SDAE) is proposed to smooth the background noise. The SDAE acts as a denoising filter for the seismic data. In the proposed algorithm, the SDAE is utilized to reduce background noise such that the onset time becomes more clear and sharp. Afterward, a hard decision with one threshold is used to detect the onset time of the event. The proposed algorithm is evaluated on both synthetic and field seismic data. As a result, the proposed algorithm outperforms the short time average/long time average (STA/LTA) and the Akaike information criterion (AIC) algorithms. The proposed algorithm accurately picks the onset time of 94.1% for 407 field seismic waveforms with a standard deviation error of 0.10 seconds. Additionally, the results indicate that the proposed algorithm can pick arrival times accurately for weak SNR seismic data with SNR higher than -14 dB.**

***Index Terms*—P-wave Arrival Time of Earthquakes, Deep Learning, and Stacked Denoising Autoencoder.**

I. Introduction

PRECISE identification of the onset time of an earthquake is necessary for correct figuring of the hypocenter location and different parameters that are utilized for building seismic catalogs. Automatic picking algorithms become essential due to the increase of seismogram database size and the existence of high background noise. In such an environment, manual picking becomes more difficult and produces false alarms for onset time detection. Accordingly, several automatic picking algorithms have been proposed to detect earthquake onset time.

Short time average/long time average (STA/LTA) and the Akaike information criterion (AIC) have been widely used algorithms for event detection and picking algorithms [1], [2]. STA/LTA has drawbacks due to difficulty configuring the parameters and the inaccuracy of onset time detection, especially with the existence of high background noise levels. For AIC, in some cases, the global minimum is incorrectly determined due to the high background noise levels. Recently, a fuzzy C-means (FCM) clustering algorithm has been proposed to detect the onset time of the event [3]. The FCM method depends on clustering the seismic data into seismic signal and seismic noise groups by obtaining the membership degree matrix. Additionally, maximal overlap discrete wavelet transform (MODWT) was proposed to detect the onset time of the event [4]. Moreover, the sliding-window cross-correlation (SCC) detection technique was proposed to detect the P-arrival time [5]. The SCC algorithm utilized a template waveform to obtain the cross-correlation coefficients of the seismic data, and then, a comparator threshold was applied to accurately detect the existence of seismic event.

Deep learning is currently an extremely powerful technique in the machine learning field [6]. It consists of multiple processing layers to learn data representations with multiple levels of abstraction [7]. An autoencoder (AE) is a deep learning approach that has distinct advantages over most other the machine learning techniques, especially when processing unlabeled data; it is based on an unsupervised technique and considered as a non-linear feature extraction method [7]. In contrast to most machine learning approaches, the input of the AE is the raw data, without the need of the input data processing [7]. AE compresses the dimensionality of the network and removes input redundancies by extracting representative features for the input data [8]. Moreover, a denoising autoencoder (DAE) is presented to be trained locally to denoise corrupted versions of their inputs [8]. After the data corruption process, using consecutive autoencoders is called a stacked denoising autoencoder (SDAE) [9]. SDAE can reconstruct the input data even with the existence of high noise levels [9]. Deep architectures can be trained in an unsupervised way independently, and fine-tuning can be accomplished using a supervisor technique that leads to superior performance [9].

Recently, deep learning techniques have been widely used in the field of remote sensing. SDAE and a deep belief network (DBN) have been utilized to classify volcano seismic events [10]. Meanwhile, a convolutional neural network (CNN) has been employed for estimating the intensity of tropical cyclones [11]. A stacked autoencoder (SAE) has been proposed as a feature extraction tool for classification of the hyperspectral data [12].

In this paper, we propose a novel methodology to accurately detect the onset time. The key idea behind our solution is to introduce SDAE for smoothing the seismic data and reducing the level of background noise. As far as we know, this is the first paper that attempts to exploit SDAE for the automatic earthquake arrival time detection. We evaluate the proposed algorithm on both synthetic and field seismic waveforms, and

Omar M. Saad is a PhD candidate in the ECE Department, E-JUST University, Alexandria 21934, Egypt, and he is with with NRIAG, Helwan 11731, e-mail: (omar.saad@ejust.edu.eg)

K. Inoue is with Kyushu University, Fukuoka 819-0395, Japan.

Ahmed Shalaby is with Benha University 13511, Egypt, and he is with E-JUST University, Alexandria 21934, Egypt.

Lotfy Samy is with the National Research Institute of Astronomy and Geophysics, Helwan 11731, Egypt.

Mohammed S. Sayed is with the ECE Department, E-JUST University, Alexandria 21934, Egypt, and he is with ECE Department, Zagazig University 44519, Egypt.

Manuscript received mm dd, yyyy; revised mm dd, yyyy.

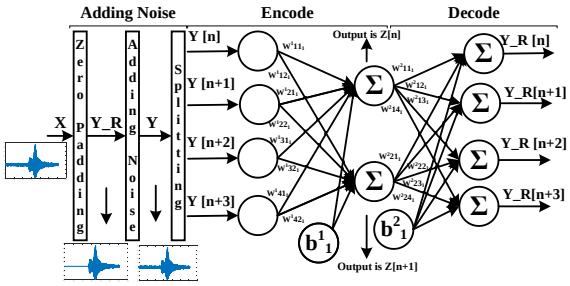

Fig. 1. First autoencoder including noise process (DAE).

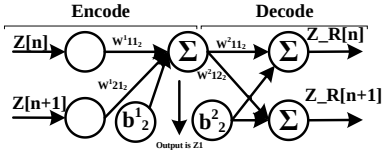

Fig. 2. Second autoencoder.

the results indicate that the proposed algorithm outperforms the previously published algorithms in the literature. This paper is organized as follows. Section II shows the proposed algorithm. Section III shows the results and analysis. Section IV concludes the work.

II. The Proposed Algorithm

The seismic data consist of two classes, the background noise, $nois$, and the seismic signal, sig. Let $X = \{nois, sig\}$ be the input seismic data. In the proposed algorithm, the background noise is smoothed such that the P-arrival time is more easily picked. Therefore, the background noise values, $nois$, are forced to be approximately zero using SDAE so that the output signal of the proposed algorithm should be approximately calculated as $Z1 = \{0, sig\}$.

To construct the SDAE, the use of two consecutive autoencoders (AEs) is proposed. There are two phases of the SDAE including the training phase and the testing phase. The training phase consists of several stages as follows:

a) The input layer is established to read and store the input data.

b) The input data are corrupted by adding noise.

c) The first AE is trained to reconstruct the input signal from the corrupted version.

d) The output signal from the encoder of the first AE is considered as the input of the second AE.

e) The second AE is trained to reconstruct its input signal.

In this paper, we divide the input signal, X, into a series of movable windows. Each window consists of four samples, $X[n : n+3]$, and the stride is one sample. Therefore, the input layer contains four perceptrons.

For noise processing, we corrupt the seismic data using the two steps shown in Fig. 1. Several seismic events are manually picked, so the background noise is well known, as is the seismic signal. First, we padded the background noise values, $nois$, with zero while keeping the seismic signal, sig, the same, $Y\_R = \{0, sig\}$. We call this signal, Y_R, as the ideal signal or noiseless signal because it has zero background noise, and this signal is considered as the input signal for the DAE. Second, the seismic signal, X, includes the background noise, and seismic signal is considered the output of the noising layer, $Y = X$. Afterward, the first AE optimizes its parameters to reconstruct the ideal signal, with zero background noise, Y_R, from the noisy signal, Y. By applying this approach, the DAE learns to reduce the background noise. DAE deals with the background noise as an undesirable signal such that its values should be forced to be approximately zero. Hence, the DAE acts as a denoising filter for the seismic data and eliminates the background noise.

The AE consists of two layers, an encoder and a decoder layer. The encoder layer extracts significant features, Z, to represent the input data. Meanwhile, the decoder layer is applied to reconstruct the input data from the encoder representation. In this paper, the encoder layer of the first AE consists of two perceptrons. For the first AE, the matrix representation of the encoder layer and the decoder layer are shown as follows:

$$\begin{bmatrix} Z[n] \\ Z[n+1] \end{bmatrix} = \begin{bmatrix} w^1 11_1 & w^1 21_1 & w^1 31_1 & w^1 41_1 \\ w^1 12_1 & w^1 22_1 & w^1 32_1 & w^1 42_1 \end{bmatrix} \begin{bmatrix} Y[n] \\ Y[n+1] \\ Y[n+2 \\ Y[n+3] \end{bmatrix} + b_1^1 \quad (1)$$

$$\begin{bmatrix} Y\_R[n] \\ Y\_R[n+1] \\ Y\_R[n+2 \\ Y\_R[n+3] \end{bmatrix} = \begin{bmatrix} w^2 11_1 & w^2 21_1 \\ w^2 12_1 & w^2 22_1 \\ w^2 13_1 & w^2 23_1 \\ w^2 14_1 & w^2 24_1 \end{bmatrix} \begin{bmatrix} z[n] \\ z[n+1] \end{bmatrix} + b_1^2 \quad (2)$$

where $W^j km_q$ and b_q^j represent the weights and biases of the AEs, respectively. j is 1 for the encoder layer or 2 for the decoder layer, while q is 1 for the first AE or 2 for the second AE. km represents the connection between the perceptrons in different layers.

Moreover, to extract significant unique features and second-order features, we added a second AE after the first AE. The output signal of the encoder layer, Z, is considered as the input of the second AE. For the second AE, the encoder layer contains one perceptron (output is $z1$), and the decoder layer reconstructs the input signal Z and stores the output signal Z_R as shown in Fig. 2. The matrix representation of the encoder and the decoder layer of the second AE are shown as follows:

$$[Z1] = \begin{bmatrix} Z[n] \\ Z[n+1] \end{bmatrix} [\, w^1 11_2 \; W^1 21_2 \,] + b_2^1 \quad (3)$$

$$\begin{bmatrix} Z\_R[n] \\ Z\_R[n+1] \end{bmatrix} = [Z1] [\, w^2 11_2 \; w^2 12_2 \,] + b_2^2 \quad (4)$$

For non-linearity purposes, we use several activation function types such as sigmoid, hyperbolic tangent and Rectified Linear Unit (ReLU). However, we use the ReLU because of its ability to reach higher performance rather than the other activation functions. The output of ReLU is defined as $max(0, ReLU_{input})$. Furthermore, for less complexity, we set the weights of the encoder layer as same as the weights of the decoder layer which called tied-weights approach [9], $W^1 km_q = W^2 km_q$. Moreover, for better training results, we use particle swarm optimization (PSO) to obtain the optimum SDAE parameters [13]. The objective of PSO is minimizing the loss function of the DAE. The PSO is used to optimize the parameters of the first and second AEs separately. The loss function, $l(f(X))$, of the first and the second AEs are $0.5 * \sum_{n=1}^{i} (Y\_R[n] - Y[n])^2$ and $0.5 * \sum_{n=1}^{i} (Z\_R[n] - Z[n])^2$, respectively [9]. For PSO, we use 50 particles to form a swarm in the design space. We use 2000 iterations, and for

each iteration, each particle updates its velocity and distance according to $v_i = v_i + rand*(P_i - d_i) + rand*(P_{global} - d_i)$ and $d_i = d_i + v_i$, respectively [13]. P_i represents the best previous position and the global best position stored in P_{global}. The velocity update formula includes some random parameters (rand) to ensure excellent coverage of the design space [13]. The PSO optimizes the parameter of the first and the second AEs independently, and then, the two AEs are stacked, and the parameters of SDAE are fine-tuned using PSO.

For the testing phase, the encoder layer of the first AE and the encoder layer of the second AE are stacked to construct SDAE as shown in Fig. 3. The seismic input signal passes through the final network topology as shown in Fig. 3. The output signal, $Z1$, is smoothed version of the input seismic data X. To detect the onset time, only one threshold, hard decision, is applied to the output signal of SDAE, $Z1$. This threshold is set for each seismic station according to its background noise level. The pseudo code of the proposed algorithm is shown as follows:

Algorithm 1 Proposed algorithm for onset time detection

Input: $X(Input\ Seismic\ Data)$
Output: $Decision(one\ for\ onset\ time\ and\ zero\ for\ background\ noise)$
for $n = 1$ **to** $i - 4$ **step** 1 **do**
  $Z[n] \leftarrow X[n] * w^1 11_1 + X[n+1] * w^1 21_1 + X[n+2] * w^1 31_1 + X[n+3] * w^1 41_1 + b_1^1$
  $Z[n+1] \leftarrow X[n] * w^1 12_1 + X[n+1] * w^1 22_1 + X[n+2] * w^1 32_1 + X[n+3] * w^1 42_1 + b_1^1$
  $Z1[n] \leftarrow Z[n] * w^1 11_2 + Z[n+1] * w^1 21_2 + b_2^1$
  if $Z1[n] >=$ threshold **then**
    $Decision \leftarrow 1$
  else
    $Decision \leftarrow 0$
  end if
  return $Decision$
end for

III. Results and Analysis

A. Synthetic Seismic Data

To create synthetic waveforms, first, we adopt the propagated algorithm proposed by Wang [14] to compute Green's functions. Second, we utilize the frequency-wavenumber (FK) method [15] to obtain several synthetic waveforms according to the determined Green's functions. The earth model used in this test is IASP91[16]. We use the open source package proposed by Wang [14] as well as a combination of depth and epicenter distance to generate several synthetic waveforms with different characteristics as follows:

Algorithm 2 Creating Synthetic Waveforms

Input: $Depth,\ and\ Epicenter\_Distance$
Output: $Synthetic\_Waveform$
for $Depth = 10$ **to** 2000 **step** 30 **do**
  for $Epicenter\_Distance = 5$ **to** 70 **step** 15 **do**
    $Using\ Wangs'\ Open\ Source\ Package$ [14] $to\ Obtain\ the\ Greens'\ Function\ and\ Synthetic\_Waveform\ according\ to\ the\ current\ Depth\ and\ Epicenter\_Distance$
    return $Synthetic\_Waveform$
  end for
end for
$Note:\ Depth\ in\ Kilometers, and\ Epicenter\_Distance\ in\ degree$

By applying this approach, we create 335 synthetic waveforms with different characteristics. Afterward, we create five models; each model consists of 335 synthetic waveforms; however, for each model, we add different Gaussian noise such

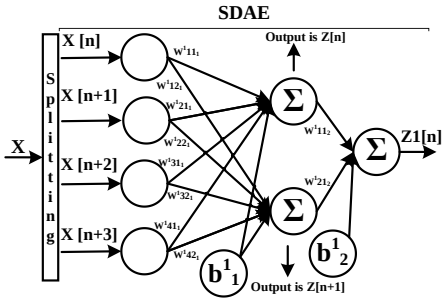

Fig. 3. Stacked Denoising Autoencoder (SDAE).

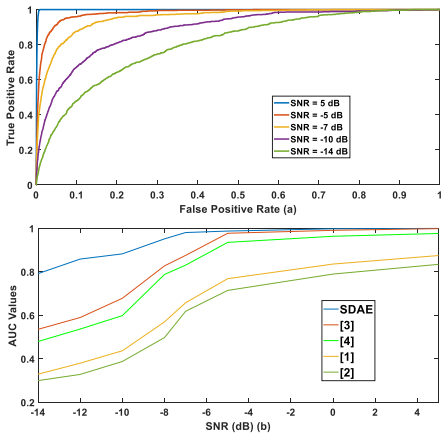

Fig. 4. (a) ROC curves for different SNRs and (b) SNR versus AUC values for [1]-[6].

that the SNR range of the five models is -14 dB to 5 dB as shown in Table I. We collect ten waveforms from each model to construct the training dataset for the SDAE. We divide each waveform to a group of windows, each window contains four samples. Afterward, the optimum SDAE parameters are obtained, and then, the proposed algorithm is tested using all the waveforms in the five models. In this paper, we define the SNR as [3], $SNR = 10log_{10}\left(\frac{\sum_n \mid sig[n] \mid^2}{\sum_n \mid nois[n] - sig[n] \mid^2}\right)$, where sig is the original signal, and $nois$ is the noisy signal [3]. The receiver operating characteristic (ROC) and area under the ROC curves (AUC) are obtained for each of the five models. AUC values reflect the accuracy of the proposed algorithm. Higher AUC values indicate higher accuracy performance of the proposed algorithm. Table I shows the accurate picking, inaccurate picking, signals pick within 0.05 seconds, and AUC values for each model. The ROC curve plots the false positive rate versus the true positive rate. Two classes are set for ROC curves, class one for the seismic signal, and class zero for the background noise. Fig. 4(a) shows the ROC curves for the proposed algorithm with different SNRs, while Fig. 4(b) shows the SNR versus the AUC values for different algorithms [1]-[4].

According to Table I and Fig. 4, the proposed algorithm outperforms the other algorithms [1]-[4]. The proposed algorithm reaches AUC values of 0.8956 and 0.7975 when the SNR results are -10 dB and -14 dB, respectively. However, the other algorithms [1]-[4] lose their ability to pick the seismic signal when the SNR goes below -8 dB, as shown in Fig. 4(b). Moreover, the majority of the signals detected by the proposed

TABLE I
AUC VALUES FOR DIFFERENT SNR SIGNALS.

Model	SNR (dB)	Signals Number	Accurate picking	Inaccurate picking	Signals pick within 0.05s	AUC values
1	5	335	335	0	335	0.9999
2	-5	335	333	2	331	0.9814
3	-7	335	327	8	322	0.9718
4	-10	335	291	44	281	0.8956
5	-14	335	266	69	225	0.7975

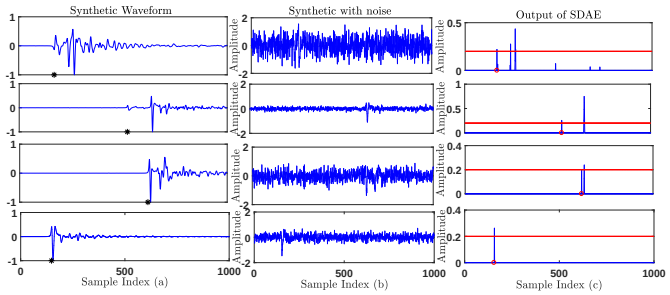

Fig. 5. (a) Input synthetic signal with different depth and epicenter distance, (b) Synthetic waveforms with different SNRs , and (c) the output signal of SDAE.

algorithm fall within 0.05 seconds, above 90%. Hence, the proposed algorithm can detect the seismic signal accurately when the SNR is higher than -14 dB. Fig. 5(a)-(c) shows four different synthetic waveforms. The depths of the four synthetic waveforms are 10, 290, 520, and 1990, respectively; meanwhile, the epicenter distances are 65, 50, 20, and 5, respectively, and the SNR results vary from -14 dB to -5 dB, as shown in Fig. 5(b). The proposed algorithm smooths the background noise, and the first sample exceeds the threshold as the P-wave arrival time of the event. Hence, the proposed algorithm detects the seismic signal accurately, as shown in Fig. 5(c). In Fig. 5, the black stars and red circles represent the onset times according to manual picking and the proposed algorithm, respectively. Meanwhile, the horizontal red line is the comparator threshold.

B. Filed Seismic Data Recorded by CNMSN

We tested the proposed algorithm using the dataset from broadband seismic stations of the Cooperative New Madrid Seismic Network (CNMSN)[5]. The dataset contains continuous data from 16 April to 2 May 2008[5]. A band-pass filter from 0.3 to 8 Hz was applied to all the waveform data[5]. Afterward, we trained the proposed algorithm on 40 events from the vertical component of the OLIL station to obtain the optimum SDAE parameters. Then, the proposed algorithm is applied on the continuous data from the vertical component of the OLIL station to detect all the P-wave arrival time. As a result, the proposed algorithm success to detect the arrival time for 142 events during this interval and 4 false alarms are detected. However, the SCC algorithm detected 151 events and 3 false alarms [5].

One of the main disadvantages of the SCC method is the window length of the template waves. In [5], the authors studied a catalog of the earthquake to calibrate the window size of the template and choose the template events with a

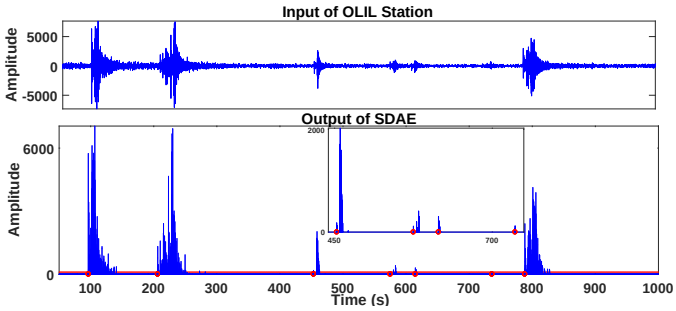

Fig. 6. Input seismic signal of OLIL Station and the output signal of SDAE.

certain amplitude. For example, the authors used a 15 seconds window size as a template event with amplitude ML of 3.3 to obtain the cross-correlation coefficients of aftershock events [5]. In the strong motion network, however, the Center for Earthquake Research and Information (CERI) was used to choose two template events with amplitudes ML of 3.3 and 3.0, respectively, with window length of 4 sec for the S-wave detection [5]. Eventually, this will discourage the application of the SCC in real-time applications and will limit the usage of the SCC algorithm in offline analysis. For example, earthuake early warning systems (EEWSs) [17] require fast detection for the P-wave arrival time. Furthermore, in [5], the main feature in the template waves is the S-wave; therefore, not all P-waves were picked correctly.

On the other hand, the proposed algorithm can support both offline and online analyses. The proposed algorithm utilizes only 4 samples to detect the arrival time. If the sample rate is 20 samples per second, the proposed algorithm can detect the arrival time within only 0.2 seconds, which is highly recommended by a real-time applications such as EEWSs [17]. A stacked denoising autoencoder (SDAE) has the ability to learn and optimize its parameters automatically to reduce the background noise such that the arrival time can be separated from the background noise. Fig. 6 shows an example for approximately 17 minutes of input data for the OLIL station and the output signal of the proposed algorithm. The background noise is well smoothed, while the onset time is picked accurately, as shown in Fig. 6. In Fig. 6, The red circles represent the onset time according to the proposed algorithm; meanwhile, the horizontal red line is the threshold comparator.

C. Field Seismic Data Recorded by ENSN

To validate and verify the robustness of the proposed algorithm, two datasets are used. The first dataset was used in [4] and contains 146 local earthquakes from Cairo region recorded by the Egyptian National Seismic Network (ENSN) and detected by the vertical component (Z-component) of the KOT, HAG, and FYM stations. The second dataset consists of 261 local earthquakes recorded by the KOT and HAG stations during 2014 and 2016.

For both datasets with a total of 407 waveforms, the earthquake magnitudes vary from 0.5ML to 5ML, while the range of SNR is -8 dB to 15 dB. Fig. 7 shows the location of these earthquakes. We use the first dataset to train the SDAE, while the second dataset is used to evaluate the performance of

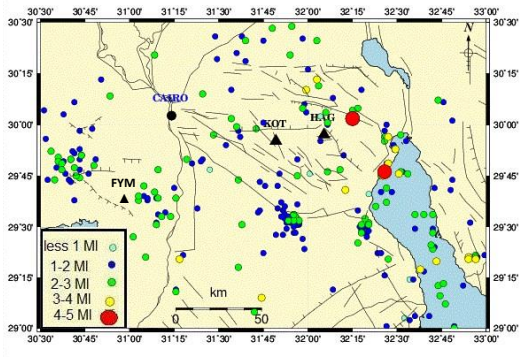

Fig. 7. Location of local earthquakes detected by the KOT, HAG, and FYM stations.

TABLE II
RESULTS COMPARISON.

Method	Accurate picking	Accuracy %	Signals pick within 0.05s	Standard Deviation(s)
[1]	310	76	178	0.110
[2]	294	72	127	0.120
[3]	359	88	318	0.110
[4]	357	87.5	246	0.130
[5]	382	93.8	347	0.101
SDAE	383	94.1	351	0.100
	The tested waveform number is 407			

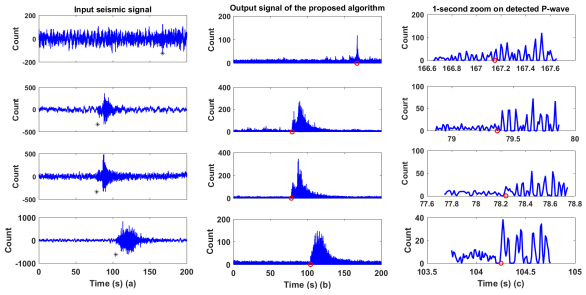

Fig. 8. (a) Input seismic signal, (b) the output signal of SDAE, and (c) 1-second zoom around P-wave arrival time.

the proposed algorithm. Table II shows the results of picking the onset time, the signals that pick with in 0.05 seconds and the standard deviation error for different algorithms[1]-[5]. The proposed algorithm picks the onset time with accuracy of 94.1% with a standard deviation error of 0.10 seconds and outperforms the other algorithms[1]-[5]. The proposed algorithm successfully picks the onset time of 351 waveforms with 91.6% accuracy in the range of 0.05 seconds from manual picking. Fig. 8(a) shows different micro/local earthquakes (4 events) with an SNR range of -8 dB to 4 dB, and the magnitude varies from 0.5ML to 2ML. Meanwhile, Fig. 8(b) shows the output signal of SDAE, and Fig. 8(c) shows a 1 second zoom around the detected onset time according to the proposed algorithm. In Fig. 8, the black stars and red circles represent the manual picking and the onset time according to the proposed algorithm, respectively.

IV. CONCLUSION

In this paper, a stacked denoising autoencoder (SDAE) is first introduced to detect the onset time. An SDAE is a deep learning technique that acts as a denoising filter for background noise. The SDAE smooths the input seismic data, and a hard decision is then applied to detect the onset time. The proposed algorithm is tested using both synthetic and field seismic data. As a result, the proposed algorithm achieves 94.1% picking accuracy of the onset time for 407 field seismic waveforms. Furthermore, ROC curves show that the proposed algorithm works perfectly with SNR values higher than $-14\ dB$. Moreover, accurate onset time is picked successfully when the proposed algorithm is tested using several micro-earthquakes with low SNR.

ACKNOWLEDGMENT

Thanks to E-JUST for the continuous support and the Egyptian Ministry of Higher Education for funding this work. Thanks to NRIAG for providing the seismic data used in this study.

REFERENCES

[1] R. V. Allen, "Automatic earthquake recognition and timing from single traces," *Bulletin of the Seismological Society of America*, vol. 68, no. 5, pp. 1521–1532, 1978.
[2] M. Leonard and B. Kennett, "Multi-component autoregressive techniques for the analysis of seismograms," *Physics of the Earth and Planetary Interiors*, vol. 113, no. 1, pp. 247–263, 1999.
[3] D. Zhu, Y. Li, and C. Zhang, "Automatic time picking for microseismic data based on a fuzzy c-means clustering algorithm," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1900–1904, 2016.
[4] A. G. Hafez, M. Rabie, and T. Kohda, "Seismic noise study for accurate p-wave arrival detection via modwt," *Computers & geosciences*, vol. 54, pp. 148–159, 2013.
[5] H. Yang, L. Zhu, and R. Chu, "Fault-plane determination of the 18 april 2008 mount carmel, illinois, earthquake by detecting and relocating aftershocks," *Bulletin of the Seismological Society of America*, vol. 99, no. 6, pp. 3413–3420, 2009.
[6] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.
[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[8] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, 2017.
[9] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. Dec, pp. 3371–3408, 2010.
[10] M. Titos, A. Bueno, L. García, and C. Benítez, "A deep neural networks approach to automatic recognition systems for volcano-seismic events," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018.
[11] R. Pradhan, R. S. Aygun, M. Maskey, R. Ramachandran, and D. J. Cecil, "Tropical cyclone intensity estimation using a deep convolutional neural network," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 692–702, 2018.
[12] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected topics in applied earth observations and remote sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.
[13] C. Sui, M. Bennamoun, and R. Togneri, "Deep feature learning for dummies: A simple auto-encoder training method using particle swarm optimisation," *Pattern Recognition Letters*, vol. 94, pp. 75–80, 2017.
[14] R. Wang, "A simple orthonormalization method for stable and efficient computation of green's functions," *Bulletin of the Seismological Society of America*, vol. 89, no. 3, pp. 733–741, 1999. [Online]. Available: avaliable Package: https://www.gfz-potsdam.de/sektion/erdbeben-und-vulkanphysik/daten-produkte-dienste/downloads-software/
[15] L. Zhu and L. A. Rivera, "A note on the dynamic and static displacements from a point source in multilayered media," *Geophysical Journal International*, vol. 148, no. 3, pp. 619–627, 2002.
[16] B. Kennett and E. Engdahl, "Traveltimes for global earthquake location and phase identification," *Geophysical Journal International*, vol. 105, no. 2, pp. 429–465, 1991.
[17] C. Peng, X. Zhu, J. Yang, B. Xue, and Y. Chen, "Development of an integrated onsite earthquake early warning system and test deployment in zhaotong, china," *Computers & geosciences*, vol. 56, pp. 170–177, 2013.