Optics EXPRESS

Improving signal-to-noise ratio of a terahertz signal using a WaveNet-based neural network

HYUNKOOK CHOI,[1,4] SANGMIN KIM,[1,4] INHEE MAENG,[2] JOO-HIUK SON,[3,5] AND HOCHONG PARK[1,6]

[1] *Department of Electronics Engineering, Kwangwoon University, Seoul 01897, Republic of Korea*
[2] *YUHS-KRIBB, Medical Convergence Research Institute, Yonsei University, Seoul 03722, Republic of Korea*
[3] *Department of Physics, University of Seoul, Seoul 02504, Republic of Korea*
[4] *These authors contributed equally*
[5] *joohiuk@uos.ac.kr*
[6] *hcpark@kw.ac.kr*

Abstract: When acquiring a terahertz signal from a time-domain spectroscopy system, the signal is degraded by measurement noise and the information embedded in the signal is distorted. For high-performing terahertz applications, this study proposes a method for enhancing such a noise-degraded terahertz signal using machine learning that is applied to the raw signal after acquisition. The proposed method learns a function that maps the degraded signal to the clean signal using a WaveNet-based neural network that performs multiple layers of dilated convolutions. It also includes learnable pre- and post-processing modules that automatically transform the time domain where the enhancement process operates. When training the neural network, a data augmentation scheme is adopted to tackle the issue of insufficient training data. The comparative evaluation confirms that the proposed method outperforms other baseline neural networks in terms of signal-to-noise ratio. The proposed method also performs significantly better than the averaging of multiple signals, thereby facilitating the procurement of an enhanced signal without increasing the measurement time.

1. Introduction

Terahertz (THz) time-domain spectroscopy (TDS) and imaging are widely used in various engineering and biomedical fields for extracting and analyzing information contained in a sample [1–6]. The THz signal is acquired from the sample by transmission or reflection, where measurement noise is inevitably introduced by signal acquisition devices. If the acquired signal is degraded by measurement noise with a low signal-to-noise ratio (SNR), embedded information is distorted, leading to the extraction of incorrect features of the sample. Therefore, an effort to obtain a THz signal with a high SNR is required for high-performing THz applications.

One solution for acquiring a high-SNR THz signal is to use a high-performing THz source and detector, or insert special equipment for reducing measurement noise [7–13]. However, this solution is not a unified methodology and should be designed and implemented for individual systems. The other solution is to convert the noisy signal to a clean signal following the signal acquisition [14–20]. This signal-based solution can be used for a given THz system without any hardware additions and modifications, regardless of its system configuration. The signal-based solution usually enhances a raw signal before feature extraction, as noise in the signal plays a significant role in providing incorrect information for the sample [16].

The simplest signal-based solution for a high SNR involves performing the same measurement multiple times and obtaining the average of the resulting signals in the time domain, thereby removing random measurement noise [14–18]. Despite its simplicity, the long measurement time required for conducting multiple measurements leads to the inability of its use for applications

#448279 https://doi.org/10.1364/OE.448279
Journal © 2022 Received 10 Nov 2021; revised 7 Jan 2022; accepted 15 Jan 2022; published 4 Feb 2022

that demand rapid signal acquisition. More sophisticated methods based on signal processing theory, such as wavelet transform and statistical modeling, have been developed for enhancing a THz signal [19,20]. Additionally, considerable research on enhancing noisy speech has been conducted in speech processing fields, and the most common approach entails the correction of degraded spectral magnitude by spectral subtraction in the frequency domain [21,22].

With advances in machine learning, it is widely used to develop methods for signal enhancement. Machine learning estimates a function from the degraded signal to the clean signal through supervised learning, which is typically conducted in the frequency domain [22,23]. The operation for spectral magnitude delivers a desirable performance, however the function from the degraded phase to the clean phase is not well learned because phase information is usually random with no distinct correlation between the degraded and clean phase. Consequently, most solutions enhance only the spectral magnitude while preserving the phase of the degraded signal, hence yielding a low enhancement performance. To resolve the limitation of the frequency-domain approach, signal enhancement with all processes conducted in the time domain has been developed [24]. However, it encountered the issue of high computational complexity due to the immense number of sample-by-sample operations in the time domain.

WaveNet, a new machine learning model specifically designed for time-domain processing of speech and audio signals, was developed recently [25]. It supports a large receptive field required for modeling long temporal dependency of the signal with moderate complexity, and has been successfully used for synthesis and modification of speech and audio signals [26,27]. WaveNet functions as an autoregressive model by feeding an output sample back to the input in the next time step, resulting in a sequential generation of output samples. The structure of WaveNet can be modified to function on a frame basis, where all output samples are determined from the input frame and independently generated without feedback; this is referred to as frame-based WaveNet [27].

In this study, as a signal-based solution for acquiring a high-SNR THz signal, we developed a method for converting the noisy THz signal to a high-SNR signal using frame-based WaveNet. We also inserted learnable pre- and post-processing modules into the network for automatically transforming the time domain where the enhancement process operates, thereby improving the enhancement performance. When training the neural network, we adopted a data augmentation scheme to solve the issue of insufficient training data. WaveNet has mostly been used for speech and audio signals, and to the best of our knowledge, this study is the first to apply the WaveNet to THz signals.

To demonstrate the superiority of the proposed method, we developed enhancement methods using other machine learning models, such as the convolutional neural network (CNN) [28,29] and conventional frame-based WaveNet, and verified that the proposed method outperforms these baseline models. To further support the contribution of this study, we also confirmed that the proposed method with only one signal measurement provides higher performance than the signal averaging method that consumes extra measurement time to acquire many signals. Because the proposed method is designed based on machine learning, it can deal with the differences in signal properties caused by different THz systems by re-training the network using a new training dataset acquired from the corresponding THz system.

In this study, as a preliminary step to verify the feasibility of the enhancement of THz signals using machine learning, we acquired the THz signals transmitted through the air and developed a network structure and its operations suitable for removing measurement noise contained in these signals. After verifying the enhancement performance for the THz signals transmitted through the air, we will extend the scope of signal enhancement to the THz signals in general THz-TDS and improve the network operations to handle the influence of test samples with different properties in our next study.

2. THz signal acquisition

Figure 1 shows the system configuration of the THz signal acquisition used in this study. We used a standard THz-TDS system based on a mode-locked Ti:sapphire laser with a pulse width of 80 fs at an 800 nm central wavelength and a repetition rate of 80 MHz. The laser beam was divided into the pump and probe beams for the generation and detection of the THz signal. The THz pulses were generated by pumping the femtosecond laser pulse on the *p*-InAs crystal. The generated THz pulses were guided and focused by the parabolic mirrors and polymethylpentene (TPX) lens onto a photoconductive detector to generate a photocurrent gated by probe pulses with a function of time delay. The THz signals were detected by acquiring photocurrents at each time delay point via a lock-in amplifier. As the time constant of a lock-in amplifier is 300 ms, the sweep time of the delay is on the order of hundreds of milliseconds per data point. It took 7 min to acquire a single 11.3 ps THz signal with 340 time samples on the time axis. The measurements were conducted in dry air, under 1% humidity condition.

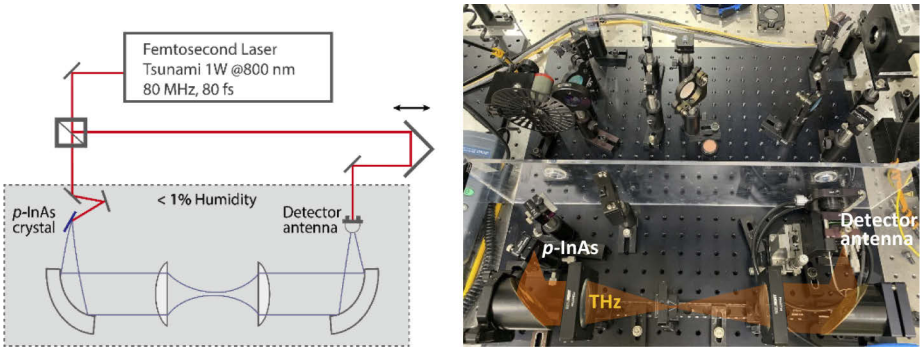

Fig. 1. System configuration of THz signal acquisition. The THz pulses are generated by pumping the femtosecond laser pulse with a width of 80 fs and a repetition rate of 80 MHz on the *p*-InAs crystal. The detector obtains the THz signals of 11.3 ps with 340 time samples on the time axis.

We acquired 40 signals and used them for machine learning and performance evaluation. Figure 2(a) shows a raw signal in the time domain with 340 time samples, denoted by $x_k(n)$, $0 \leq n < 340$, where $1 \leq k \leq 40$ is a measurement index, and the inset presents the noise floor of three different signals. To run supervised learning for signal enhancement, we require an ideal noise-free signal with infinite SNR that is used as a target signal for training. However, it is impossible to acquire such a signal in practice using the same THz system. Instead, assuming that most of the noise is removed by averaging many noisy signals, we use the mean of the signals provided for training as the target signal, which is denoted by $\bar{x}(n)$. Figure 2(b) shows the spectral magnitudes of $x_k(n)$ and $\bar{x}(n)$ averaged over 38 training signals, which were obtained by a discrete Fourier transform. We can see that noise characteristics vary with $x_k(n)$ and $\bar{x}(n)$ has a peak dynamic range of approximately 73 dB. Then, the goal of enhancement in this study is to determine a function that maps $x_k(n)$ to $\bar{x}(n)$ for all k.

Because $\bar{x}(n)$ is assumed to be a noise-free signal, $e_k(n) = [\bar{x}(n) - x_k(n)]$ corresponds to the noise contained in $x_k(n)$ for each k. Then, we measure the quality of $x_k(n)$ in the time domain using the SNR defined in Eq. (1), which corresponds to the ratio of the power of $\bar{x}(n)$ to the power of $e_k(n)$.

$$SNR_k(\text{dB}) = 10\log_{10}\frac{\sum_n |\bar{x}(n)|^2}{\sum_n |e_k(n)|^2} \quad (1)$$

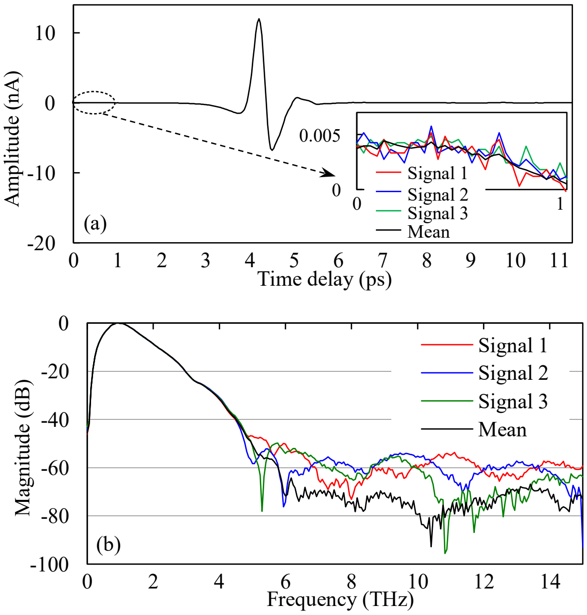

Fig. 2. Examples of THz signals. (a) raw signal of 11.3 ps with 340 time samples in the time domain with the inset showing the noise floor of three different signals and the mean of 38 signals; (b) spectral magnitudes of three different signals and the mean of 38 signals.

The average SNR, computed by averaging SNR_k over k, is 29.82 dB, which corresponds to the baseline quality of the raw signals before enhancement. Note that the SNR in Eq. (1) is different from the peak dynamic range, which is also often referred to as an SNR in THz field [17].

3. Proposed enhancement method

3.1. WaveNet structure

There are various machine learning models, such as the neural network (NN), support vector machine, and decision tree [29]. We selected the NN because it is suitable for conducting signal enhancement and various structural variations have been developed for it. The feedforward NN consists of multiple fully-connected layers, where the neurons in the adjacent layers are all connected. The structure of the NN can be modified to analyze the local properties of an input using convolution with a small filter size, resulting in the CNN [29]. The CNN usually includes a pooling operation to extract only key local features.

In this study, we use the WaveNet model that is a variant of NN designed for conducting signal processing in the time domain, resulting in the elimination of problems that occur in a frequency-domain approach. Specifically, we use frame-based WaveNet, instead of the original WaveNet operating in a recursive fashion, because a non-causal frame-to-sample operation is more suitable for enhancing 340-sample signal than generating samples one-by-one in a recursive manner.

The structure of WaveNet is primarily the same as that of a CNN with the exception that the filter is applied to the input area larger than the filter size by skipping input samples. Figure 3 shows the structure of a frame-based non-causal WaveNet with five layers that serves as the framework of the proposed network. In the input layer, each neuron is associated with one sample out of 340 input samples. Each neuron in the output layer corresponds to one output sample, and its value is computed by a convolution of three neurons in the previous hidden layer that are eight positions apart; this operation is called a dilated convolution with a filter size of 3 and a dilation factor, D, of 8. The same connection between layers is repeated with the same filter size, while decreasing the dilation factor by half up to the input layer. All neurons use the gated activation with hyperbolic tangent and sigmoid functions as in [25]; no residual connection is used. This is how a stack with five layers, as shown in Fig. 3, is constructed, where one output sample is determined by 31 input samples; this structure is said to have the receptive field of 31 samples. As a result, we obtain a frame-to-sample function in the form of layered dilated convolutions. The next output sample is also determined by a 31-sample input frame, centered at the output time position. Consequently, when all the input samples are given, all output samples are computed in parallel using different input frames of 31 samples. In the training stage, we learn all the convolution filter coefficients and obtain a fixed function that maps the 31-sample frame to one output sample in all the time positions.

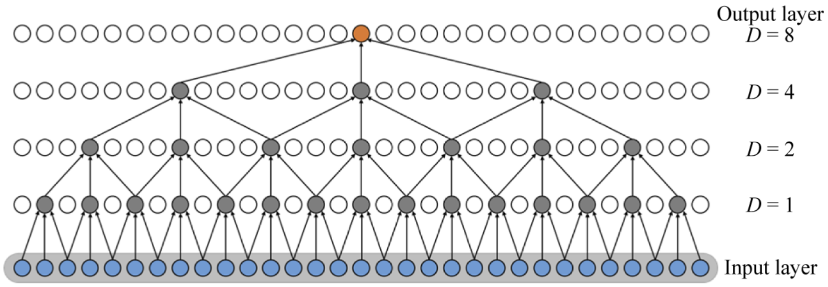

Fig. 3. Structure of frame-based non-causal WaveNet. It conducts convolution with a filter size of 3 and a dilation factor of 1, 2, 4, or 8 in each of five layers and serves as the framework of the proposed network. The receptive field of this structure is 31 samples.

The superiority of WaveNet, compared to the CNN, stems mainly from a large receptive field with a small filter size, made possible by the dilated convolutions [25]. The network with a large receptive field has the potential for a better performance because it can model long temporal dependency of the signal using input samples in a long time period. When more layers are added to the top of the stack in Fig. 3 with increasing dilation factors, the receptive field of WaveNet increases. However, as the network size increases with more layers, the risk of poor training also increases, especially when the size of the training dataset is small.

3.2. Pre- and post-processing

As shown in Fig. 2(b), most of the signal power lies in the low frequency band. Then, as the network training continues, the enhancement function is biased to remove noise in the low band with less noise removal in the high band, because noise removal in the high-power band is more helpful in reducing the cost function that guides the direction of network learning. To solve the problem of low-band-focused learning, a pre-emphasis on the high band is usually applied to the signal before training [24]. Apart from pre-emphasis on the high band, it is also necessary to increase the effective receptive field of WaveNet without increasing its size, because signal enhancement using input samples in a longer time period can yield better performance.

Although the two tasks of high-band emphasis and receptive field increase are not related, we design a novel method to achieve the effects of both the tasks through a single process. We apply a learnable down-sampling operation to the input through the NN. Subsequently, the WaveNet operates on the down-sampled time axis with lower temporal resolution, which effectively increases its receptive field when viewed from the original time axis. In addition, the down-sampling operation causes spectrum aliasing that makes the spectral envelope of signal flatter, similar to the effect of pre-emphasizing the high band. The down-sampling is paired with the learnable up-sampling including a function for aliasing removal at the final stage of enhancement.

The pre-processing used in the proposed method is different from the conventional down-sampling in the way the coefficients are determined. The conventional down-sampling conducts a fixed operation with pre-defined coefficients without considering the enhancement operation, whereas the pre-processing uses the coefficients that are learned in connection with the enhancement operation. Therefore, the pre-processing automatically transforms the time domain such that the subsequent WaveNet module performs the optimal enhancement.

3.3. Proposed network for signal enhancement

We design an enhancement network that consists of a frame-based WaveNet module and pre- and post-processing modules, as shown in Fig. 4. The pre-processing module, serving as down-sampling, consists of one CNN layer of convolution and max pooling. It inputs the noisy signal $x_k(n)$ of 340 samples and outputs the pre-processed signal $x_k{}'(n)$ of 170 samples. In all convolution and pooling boxes in all figures, the numbers in parenthesis indicate the filter size and the number of output channels. For example, the convolution box in the pre-processing module conducts the convolution with a filter size of 3 and outputs a one-channel signal. The convolution uses a stride of 1 and max pooling uses a stride of 2, which then implements the down-sampling by 2.

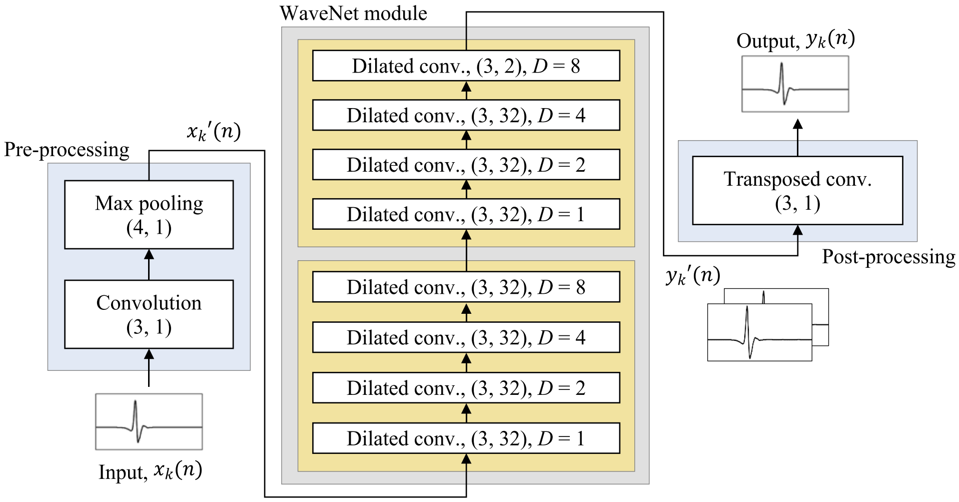
WaveNet module
Dilated conv., (3, 2), D = 8
Dilated conv., (3, 32), D = 4
Dilated conv., (3, 32), D = 2
Dilated conv., (3, 32), D = 1
Dilated conv., (3, 32), D = 8
Dilated conv., (3, 32), D = 4
Dilated conv., (3, 32), D = 2
Dilated conv., (3, 32), D = 1
Pre-processing
$x_k{}'(n)$
Max pooling (4, 1)
Convolution (3, 1)
Input, $x_k(n)$
Output, $y_k(n)$
Transposed conv. (3, 1)
Post-processing
$y_k{}'(n)$

Fig. 4. Overall structure of the proposed enhancement network. It consists of pre-processing, WaveNet, and post-processing modules, where the WaveNet module contains two stacks of frame-based WaveNet. The pre- and post-processing modules conduct learnable down- and up-sampling operations to increase the receptive field of the overall network and emphasize the high band.

We input $x'_k(n)$ of 170 samples to the WaveNet module and compute its output $y_k{}'(n)$ of 170 samples, which leads the WaveNet module to run on the down-sampled time axis. Here, the WaveNet module consists of two WaveNet stacks shown in Fig. 3, where each dilated convolution outputs a 32-channel signal, except for the last one with a two-channel output. We then input the two-channel signal $y_k{}'(n)$ to the post-processing module, serving as an up-sampling with aliasing removal, and obtain the final enhanced signal $y_k(n)$ of 340 samples through one CNN layer of transposed convolution with a stride of 2. No biases are used in all modules, and the pre-processing, WaveNet, and post-processing modules have 7, 18720, and 6 parameters, respectively. When training the network in Fig. 4, we run a joint optimization for all pre-processing, post-processing, and WaveNet modules in each learning step. In this way, we can learn the down- and up-sampling operations in connection with the WaveNet function, such that the optimal enhancement is obtained.

The receptive field of the WaveNet module in Fig. 4 is 61 samples; it almost doubled compared to that of WaveNet in Fig. 3 because of the two-stack structure. The effective receptive field of the overall network in Fig. 4 becomes 124 samples. In other words, owing to the down-sampling in the pre-processing module, the receptive field almost doubles without increasing the size of the WaveNet module. Accordingly, each output sample in $y_k(n)$ is computed from 124 samples in $x_k(n)$, centered at the output time position. To get a similar receptive field without the pre- and post-processing modules, we have to insert two more stacks to the WaveNet module, which then doubles the WaveNet size

4. Performance evaluation

4.1. Experimental setup

To verify the superiority of the proposed method, we included two additional NN models as baseline models for performance evaluation. One baseline model is a CNN autoencoder with a bottleneck structure with seven layers, as shown in Fig. 5. All convolutions and transposed convolutions use a stride of 2, and the pooling operation is not used in any layers. The number of signal channels was set the same as in the proposed WaveNet. Each of the 32 channels in the output of the third convolution has fewer samples than the input; therefore, it learns to represent core information on the input, with different aspects for different channels, that is essential for reconstructing the desired output. Then, unnecessary information such as the noise component is eliminated as the input passes through the network and the desired signal enhancement is conducted. The other baseline model is the conventional frame-based WaveNet without pre- and post-processing modules. This baseline is identical to the WaveNet module in Fig. 4, except for the last dilated convolution layer. As there is no post-processing for the two-to-one channel mapping, the final dilated convolution layer generates a one-channel output that corresponds to the final enhanced signal.

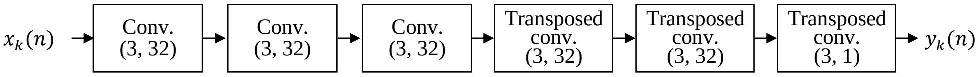

Fig. 5. Baseline CNN for performance comparison. It has an autoencoder structure with bottleneck. The three convolution layers encode the input to represent core information for noise elimination, and the three transposed convolution layers reconstruct the enhanced signal from the core information.

We only have 40 signals $x_k(n)$, which are insufficient to conduct a general K-fold cross validation for performance evaluation [29]. Hence, we implemented a leave-one-out cross validation (LOOCV), which is widely used in performance evaluation for machine learning with insufficient data [29]. In each trial for the LOOCV, we used 38 signals for training, one for validation, and one for testing. For each network in the evaluation, we ran 40 independent trials with different training signals and obtained 40 learned models, each using 38 training signals. We then evaluated the performance of each model using one testing signal associated with the trial. In this manner, every $x_k(n)$ participated once in the testing. Finally, we computed the average performance across the 40 models as the final performance of the given network.

All networks in the evaluation were trained using the same methods for a fair comparison. They used the He initialization [30], stochastic gradient descent (SGD) method with a batch size of one [29], Adam optimizer [31], mean-squared-error cost function with no regularization terms, and the learning rate of 0.005. To prevent overfitting, early stopping was used with a patience time of 100 epochs. All networks were implemented and trained using a TensorFlow platform. We confirmed that the learning process of each network does not contain any abnormal phenomena and shows a learning curve of typical shape observed in normal machine learning.

4.2. Preliminary performance

Table 1 shows the average SNR of enhanced signals across all 40 trials in the LOOCV for the evaluated networks, including the SNR of unprocessed raw signals, where the SNR of each signal was computed using Eq. (1). The proposed WaveNet provides the highest SNR, and the baseline WaveNet is better than the baseline CNN. These results confirm that the pre- and post-processing modules in the proposed WaveNet, designed for emphasizing the high band and increasing the receptive field, improve the performance, compared to the baseline WaveNet, despite the fact that both networks use the same WaveNet module. For the significance test for evaluation results, we conducted the *t*-test over 40 trials in the LOOCV and confirmed that the proposed WaveNet provides significant improvement with the *p*-value less than 0.01, compared to both baseline models.

Table 1. Preliminary Performance for Various Networks

Network	Avg. SNR (dB)
No enhancement (raw signal)	29.82
Baseline CNN	43.38
Baseline WaveNet	51.32
Proposed WaveNet	53.93

Let $y_k(n)$ be the enhanced signal from $x_k(n)$. Figure 6 shows the spectral magnitude of $y_k(n)$ by different methods for each of three $x_k(n)$ shown in Fig. 2(b), along with the spectral magnitudes of $x_k(n)$ and its target $\bar{x}(n)$ made by averaging 38 training signals. Owing to the enhancement process, $y_k(n)$ has a spectral magnitude closer to the target than $x_k(n)$ with every method. Even with the proposed method, however, the high-band noise level of $y_k(n)$ is definitely higher than that of the target, which implies that the enhancement performance has not yet reached the desired level. Therefore, a better training scheme is required, while maintaining the same network structure, in order to further decrease the high-band noise level after enhancement.

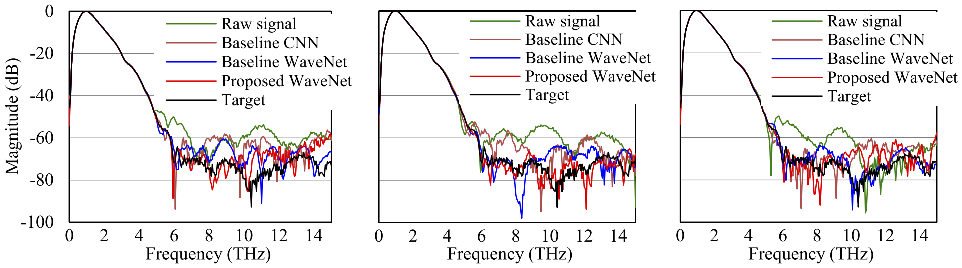

Fig. 6. Spectral magnitude of enhanced signal by different methods for three different signals. For each signal, the high-band noise level of enhanced signal is higher than that of the target, even using the proposed WaveNet. These results mean that the enhancement performance has not yet reached the desired level.

4.3. Data augmentation

When 38 training signals are given in each trial of the LOOCV, the training performance is poor owing to the amount of information in 38 training signals being insufficient to represent general properties of a target model, hence causing overfitting of learning to these signals. To improve the training performance by solving this problem, we developed data augmentation method that artificially increases the number of training signals with possible enhancement in generalization, without extra signal measurements. We used three methods of data augmentation and compared their performances. First, we shifted each training signal on the time axis to generate new signals. Because the signal samples are almost zero, except in the mid-region as shown in Fig. 2(a), we shifted the signal by a small distance without modifying its characteristics. For each training signal, we obtained 70 new time-shifted signals with time offsets between -20 and 50, and the total number of training signals after data augmentation became $38 \times 71 = 2698$ for each of the 40 trials in the LOOCV.

The second method of data augmentation was the generation of new signals by merging two different signals via averaging. Because all signals contain the same information with different measurement noise, the signal merge can generate new signals that are likely to be acquired using the same THz system, although they have lower noise level than the raw signals because of the averaging effect. Considering all cases of signal pairs from 38 training signals, we generated $(38 \times 37)/2 = 703$ new signals, and the total number of training signals became $38 + 703 = 741$ for each of the 40 trials in the LOOCV. The third method was a combination of the two methods of time shift and signal merge. In this case, the final number of training signals became $741 \times 71 = 52611$ for each trial in the LOOCV.

We trained the network using each of the three augmented training datasets and compared their performances. When the time-shift data augmentation and the combined data augmentation were used, the batch size in the SGD changed to 16, in accordance with the large number of training signals. Data augmentation was applied only to the training signals, and all evaluations were performed using the original raw signals.

4.4. Performance of signal enhancement

The average SNR of enhanced signals per data augmentation method is shown in Table 2. For each network, data augmentation improves the enhancement performance, compared to no data augmentation, and the degree of SNR improvement heavily depends on the data augmentation method. The signal merge functions better than the time shift because the former generated signals with new shapes, whereas the latter only changed time positions. The combination of the two augmentation methods yields the best performance, except for the CNN, as it provides the largest number of training signals with varying shapes and positions. For each data augmentation method, the proposed WaveNet continues to provide a higher SNR than both baseline networks with the p-value less than 0.01 in the t-test, and the baseline WaveNet is better than the CNN. Finally, we obtain the average SNR of 71.49 dB when using the proposed WaveNet along with combined data augmentation, which provides an increase of 41.67 dB compared to the unprocessed raw signals with an SNR of 29.82 dB.

Figure 7 shows the spectral magnitude of $y_k(n)$ by different enhancement methods when the combined data augmentation is used for the same $x_k(n)$ shown in Fig. 6. With the baseline WaveNet or the proposed WaveNet, $y_k(n)$ and its target $\bar{x}(n)$ have almost the same high-band noise level, which proves the ability of data augmentation for generalized training. As a result, the proposed method increases the peak dynamic range of acquired signal to the target level. We also confirmed that $y_k(n)$ by the proposed method and its target have almost the same phase over the entire frequency range. Even with data augmentation, however, the baseline CNN cannot provide the desired level of enhancement, which confirms that the WaveNet has a better ability to enhance the THz signals than the CNN. In conclusion, the enhanced signal by the proposed method has

Optics EXPRESS

Table 2. Performance for Various Networks and Data Augmentation Methods

Data augmentation	Network	Avg. SNR (dB)
Time shift	Baseline CNN	44.64
	Baseline WaveNet	58.41
	Proposed WaveNet	66.43
Signal merge	Baseline CNN	55.14
	Baseline WaveNet	64.84
	Proposed WaveNet	68.91
Combination	Baseline CNN	49.86
	Baseline WaveNet	66.98
	Proposed WaveNet	71.49

similar spectral characteristics and noise level to the target that was made by averaging 38 signals. Further, the enhancement process does not introduce any spectral distortion in both magnitude and phase. Therefore, we can replace the average of multiple signals, which is commonly used as a high-SNR signal, with the enhanced signal without causing any fundamental differences in THz applications.

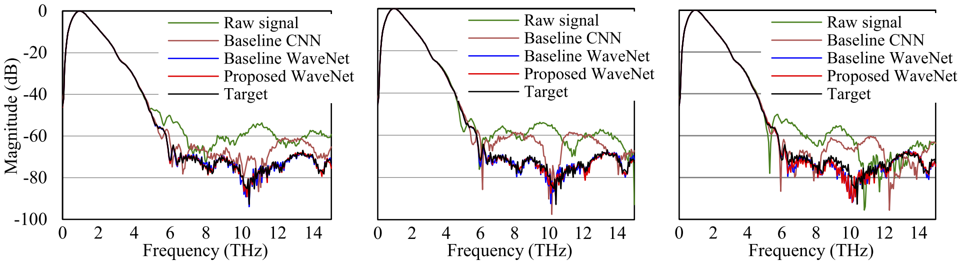

Fig. 7. Spectral magnitude of enhanced signal by different methods when the combined data augmentation is used. For each signal, the proposed WaveNet reduces the high-band noise level to the target noise level and the enhanced signal has the same peak dynamic range as the target signal.

To further increase the performance, we conducted the experiments on signal enhancement with more than one input signal for the proposed WaveNet architecture. We acquired THz signals K times and conducted the signal enhancement through simultaneous application of K signals to the proposed WaveNet in a K-channel format. To handle the K-channel input, the pre-processing module in Fig. 4 changed, as shown in Fig. 8. The K-channel input with 340 samples per channel is converted to a one-channel signal of 170 samples through convolution and max pooling. Beyond this stage, all operations are the same as before, with the output in Fig. 8 being applied to the WaveNet module in Fig. 4. This is how K noisy signals are applied to the proposed WaveNet and one enhanced signal is generated. This structure has the potential to improve the enhancement performance using the correlation among K signals, as well as individual signal characteristics. Because the number of input samples increases in the pre-processing module, it is necessary to increase the network size to manipulate more input information. However, to analyze the effect of the number of input signals, we use the same pre-processing architecture.

Figure 9 shows the SNR as a function of the number of input channels, where the proposed WaveNet was independently trained for each number of channels, using one of the three data augmentation methods. The highest SNR of 72.2 dB is achieved with four input channels using

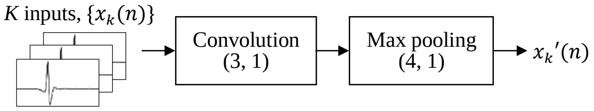

Fig. 8. Structure of the pre-processing module for K-channel input in the proposed network. The convolution layer converts the K-channel input into one-channel signal, which is then applied to the max pooling.

combined data augmentation, whereas the use of more than four input channels yields a lower SNR. The lower SNR with more than four input channels is partly due to the small pre-processing size with respect to the increased input channels. These experiments confirms that we can further increase the SNR by spending more measurement time to acquire three or four signals.

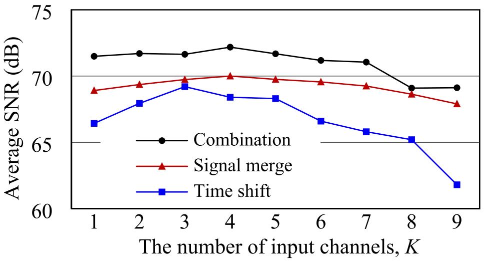

Fig. 9. Performance of multi-channel signal enhancement for the proposed network. The highest SNR of 72.2 dB is achieved when using four input channels and combined data augmentation.

4.5. Comparison with averaging method

As a final step in performance evaluation, we compared the performance of the proposed method and the averaging method that is the most common method for obtaining high-SNR signal. We measured the SNR of the averaging method with N measurements as follows. We selected N signals at random out of 40 signals $x_k(n)$ and measured the SNR of the averaged signal. We repeated this process 40 times for each N and computed the average SNR over 40 trials. Figure 10 shows the average SNR as a function of N, where the blue and red lines represent the SNR values obtained by the proposed method with and without data augmentation, respectively. For the averaging method, a monotonic increase of SNR is evident as N increases. When data augmentation is not used, the averaging method with 33 measurements provides the same SNR as the proposed method with only one signal measurement. In this case, the measurement time for the averaging method is $33 \times 7 = 231$ min, whereas the measurement time for the proposed method is 7 min. When data augmentation is used, the proposed method performs better than the averaging of any number of signals.

In summary, the performance evaluation corroborates that the proposed WaveNet-based neural network with one signal measurement improves the SNR and reduces the high-band noise to a level that can be achieved by averaging many signals. As a result, the proposed method significantly reduces the measurement time, compared to the averaging method, when obtaining high-SNR THz signals. The quality of THz signal is a key requirement for high-performing THz applications regardless of the type of task, and this study contributes to the THz field by providing a new way to obtain high-SNR signals without increasing the measurement time.

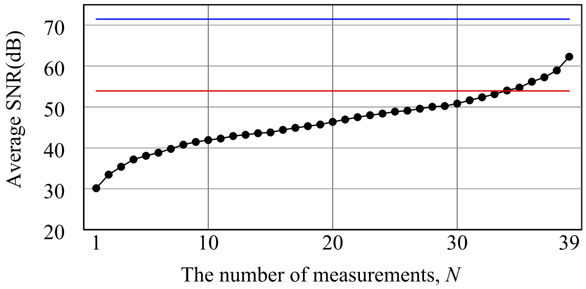

Fig. 10. Performance of signal averaging method. The SNR increases monotonically with the number of signal measurements. The blue and red lines represent the SNR values obtained by the proposed method with and without data augmentation, respectively. The proposed method without data augmentation provides similar performance to the averaging of 33 signals.

5. Conclusion

In this study, we proposed a method to enhance noise-degraded THz signals using a WaveNet-based neural network that is required for high-performing THz applications. By applying a series of dilated convolutions to the input samples in the time domain, we estimated an enhancement function from the noisy signal to the high-SNR signal. We also inserted pre- and post-processing modules with a learnable filter to transform the time domain for better enhancement. Through comparative performance evaluation, we verified that the proposed WaveNet outperforms the conventional frame-based WaveNet and CNN. When multi-channel input is applied to the proposed WaveNet, input signals up to four yield a higher SNR. Finally, we confirmed that the proposed WaveNet provides a higher SNR than the averaging of signals after multiple measurements. Despite acquiring only 40 signals for machine learning, we obtained an enhancement network with satisfactory performance, which implies that we will be able to achieve the same performance using this network for new THz systems without the heavy burden of signal acquisition to prepare new training datasets. This study verifies the feasibility of improving the SNR of THz signal using the proposed network architecture. In our future research, we intend to investigate the operations of the proposed network for the THz signals in general THz-TDS and imaging set-ups.

Funding. National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2021R1F1A1059233); Kwangwoon University (2021 Excellent researcher support project); University of Seoul (2018 University of Seoul Research Grant (JHS)).

Disclosures. The authors declare no conflicts of interest.

Data availability. No data were generated or analyzed in the presented research.

References

1. Y. C. Sim, J. Y. Park, K. M. Ahn, C. Park, and J.-H. Son, "Terahertz imaging of excised oral cancer at frozen temperature," Biomed. Opt. Express **4**(8), 1413–1421 (2013).
2. K. Kim, D.-G. Lee, W.-G. Ham, J. Ku, S.-H. Lee, C.-B. Ahn, J.-H. Son, and H. Park, "Adaptive compressed sensing for the fast terahertz reflection tomography," IEEE J. Biomed. Health Inform. **17**(4), 806–812 (2013).
3. J.-H. Son (Ed.), *Terahertz biomedical science and technology* (CRC Press, 2014).
4. H. Cheon, H. J. Yang, S. H. Lee, Y. A. Kim, and J.-H. Son, "Terahertz molecular resonance of cancer DNA," Sci. Rep. **6**(1), 1–10 (2016).
5. M. Naftaly, N. Vieweg, and A. Deninger, "Industrial applications of terahertz sensing: state of play," Sensors **19**(19), 4203 (2019).
6. J.-H. Son, S. J. Oh, and H. Cheon, "Potential clinical applications of terahertz radiation," J. Appl. Phys. (Melville, NY, U. S.) **125**(19), 190901 (2019).
7. T. Kataoka, K. Kajikawa, J. Kitagawa, Y. Kadoya, and Y. Takemura, "Improved sensitivity of terahertz detection by GaAs photoconductive antennas excited at 1560 nm," Appl. Phys. Lett. **97**(20), 201110 (2010).

8. L. Hou, W. Shi, and S. Chen, "Noise analysis and optimization of terahertz photoconductive emitters," IEEE J. Sel. Top. Quantum Electron. **19**(1), 8401305 (2013).
9. K. Murate, Y. Taira, S. R. Tripathi, S. Hayashi, K. Nawata, H. Minamide, and K. Kawase, "A high dynamic range and spectrally flat terahertz spectrometer based on optical parametric processes in LiNbO$_3$," IEEE Trans. THz Sci. Technol. **4**(4), 523–526 (2014).
10. X. Qiao, X. Zhang, J. Ren, D. Zhang, G. Cao, and L. Li, "Mean estimation empirical mode decomposition method for terahertz time-domain spectroscopy de-noising," Appl. Opt. **56**(25), 7138–7145 (2017).
11. S. Pang, Y. Zeng, Q. Yang, B. Deng, H. Wang, and Y. Qin, "Improvement in SNR by adaptive range gates for RCS measurements in the THz region," Electronics **8**(7), 805 (2019).
12. Y. Peng, C. Shi, Y. Zhu, M. Gu, and S. Zhuang, "Terahertz spectroscopy in biomedical field: a review on signal-to-noise ratio improvement," PhotoniX **1**(1), 12–18 (2020).
13. Z. Zhaohui, L. Yongli, and Z. Xinyong, "Design of an adaptive lock-in amplifier for the terahertz system," J. Phys.: Conf. Series **1865**(2), 022010 (2021).
14. B. M. Fischer, M. Hoffmann, and P. U. Jepsen, "Dynamic range and numerical error propagation in terahertz time-domain spectroscopy," *Optical Terahertz Sci. Tech.*, TuD1 (2005).
15. M. Naftaly and R. Dudley, "Methodologies for determining the dynamic ranges and signal-to-noise ratios of terahertz time-domain spectrometers," Opt. Lett. **34**(8), 1213–1215 (2009).
16. W. Withayachumnankul and M. Naftaly, "Fundamentals of measurement in terahertz time-domain spectroscopy," J. Infrared, Millimeter, Terahertz Waves **35**(8), 610–637 (2014).
17. N. Vieweg, F. Rettich, A. Deninger, H. Roehle, R. Dietz, T. Göbel, and M. Schell, "Terahertz-time domain spectrometer with 90 dB peak dynamic range," J. Infrared Millim. Terahertz Waves **35**(10), 823–832 (2014).
18. J. Neu and C. A. Schmuttenmaer, "Tutorial: an introduction to terahertz time domain spectroscopy (THz-TDS)," J. Appl. Phys. **124**(23), 231101 (2018).
19. X. Chen and E. Pickwell-MacPherson, "Signal denoising algorithm for terahertz imaging and spectroscopy," in *Proceedings of International Conference on Infrared, Millimeter, and Terahertz Waves* (2019).
20. Z. Zhang, Y. Lu, C. Lv, Q. Mao, S. Wang, and S. Yan, "Restoration of integrated circuit terahertz image based on wavelet denoising technique and the point spread function model," Opt. Lasers Eng. **138**, 106413 (2021).
21. N. Upadhyay and A. Karmakar, "Speech enhancement using spectral subtraction-type algorithms: a comparison and simulation study," Procedia Comput. Sci. **54**, 574–584 (2015).
22. J. Umamaheswari and A. Akila, "Improving speech recognition performance using spectral subtraction with artificial neural network," Int. J. Advan. Studies Sci. Research **3**(11), 1 (2018).
23. C. Donahue, B. Li, and R. Prabhavalkar, "Exploring speech enhancement with generative adversarial networks for robust speech recognition," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, 2018), pp. 5024–5028.
24. S. Pascual, A. Bonafonte, and J. Serra, "SEGAN: speech enhancement generative adversarial network," arXiv preprint, arXiv:1703.09452 (2017).
25. A. V. D. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "WaveNet: a generative model for raw audio," arXiv preprint, arXiv:1609.03499 (2016).
26. J. Engel, C. Resnick, A. Roberts, S. Dieleman, M. Norouzi, D. Eck, and K. Simonyan, "Neural audio synthesis of musical notes with WaveNet autoencoders," in *Proceedings of International Conference on Machine Learning* (2017), pp. 1068–1077.
27. D. Rethage, J. Pons, and X. Serra, "A wavenet for speech denoising," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, 2018), pp. 5069–5073.
28. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**(7553), 436–444 (2015).
29. H. Park and J.-H. Son, "Machine learning techniques for THz imaging and time-domain spectroscopy," Sensors **21**(4), 1186 (2021).
30. K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of IEEE International Conference on Computer Vision* (IEEE, 2015), pp. 1026–1034.
31. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv preprint, arXiv:1412.6980 (2014).