End-to-End Full-Waveform Echo Decomposition Based on Self-Attention Classification and U-Net Decomposition

Gangping Liu [ID], *Member, IEEE*, and Jun Ke [ID], *Member, IEEE*

Abstract—**Different from conventional decomposition methods, which utilize several steps to obtain the final result, a self-attention-based neural network, Attention Full-waveform Decomposition Network (AFD-Net), is discussed in this article for end-to-end full-waveform LiDAR signal decomposition. In existing LiDAR waveform decomposition methods, complicated functional models are used to fit echo components. Thus, the echo decomposition problem can be translated into a function approximation task. Recent studies present great progress in estimating the parameters of fitting models, hence, in the final decomposition results. However, the shape of received echo components is always irregular. None of the parametric functional models can fit the received echo components perfectly, which leads to unavoidable errors in the initial step of echo decomposition. In this article, we propose an end-to-end network AFD-Net to solve the echo decomposition problem without assuming any parametric functional models. AFD-Net consists of two modules: 1) the classification module and 2) the decomposition module. The former module is used to determine the number of echo components in a received waveform. Then, the decomposition module is used to output the echo components. By experiments, we have a classification accuracy of 96% using the first module. The average R^2 coefficient for the decomposed echo components is 0.9799. In addition, there are no public datasets for the waveform decomposition task available. Thus, another contribution of our work is to develop a tool to generate synthetic full-waveform LiDAR signals, which can help researchers to construct their own dataset for related works.**

Index Terms—**End-to-end neural network, full-waveform datasets, full-waveform LiDAR, self-attention, waveform decomposition.**

I. INTRODUCTION

LIGHT detection and ranging (LiDAR) is a wildly used remote sensing (RS) technique, which could provide accurate measurements between a laser source and reflective targets [1], [2], [3]. However, most time-of-flight LiDARs only

provide the coordinates of targets through discrete echo points. Later full-waveform (FW) LiDAR systems have been developed. Such a system can record the whole back-scattered signals at a temporal resolution of nanosecond or subnanosecond. Thus, much more detail about the reflective targets can be extracted from a waveform signal [4], [5], [6]. FW LiDAR has great potential in various applications such as topography [2], [7], [8], characterizing forest canopy structure [9], [10], [11], estimating terrain slope [12], and objection classification [13], [14], [15].

To analyze FW signals, usually echo decomposition is required first. Researchers have proposed different methods to solve the echo decomposition problem [16], [17]. However, challenges still remain in increasing decomposition speed and accuracy, especially for multiple return echoes. In [16], Hofton et al. decomposed a waveform signal into several Gaussian components, assuming that each Gaussian component represents a reflective surface. In this work, the received waveform is first preprocessed for denoising. Then, the inflection points of the preprocessed signal are used to estimate the number of the initial half-widths and the positions of Gaussian components. After that further parameter estimations are obtained by solving an optimization problem using the Levenburg–Marquardt (LM) [18] method. This work was a pioneer work for FW LiDAR waveform decomposition.

Following this work, more decomposition methods have been developed. Generally, there are the following three basic steps:

1) selecting a fitting model for echo components;
2) preprocessing the received raw FW data for denoising;
3) parameter estimations using optimization methods.

For the first step, existing methods usually assume functions such as Gaussian, generalized Gaussian, Weibull, and Nakagami to model the echo components in an FW signal [19]. For the popular Gaussian fitting model, it has two assumptions: 1) the emitted FW LiDAR pulse has a Gaussian shape; 2) the returned FW LiDAR signal is a mixed Gaussian waveform, in which the mean and the standard deviation of a Gaussian component are determined by the emitted laser pulse and the reflective object surface. However, these assumptions are not always true. First, the Gaussian function cannot describe the emitted laser pulse perfectly. Moreover, after transmitting over a long distance, the signal can deviate from Gaussian. For example, one of the properties of Gaussian waves is symmetry, but a multipath reflection will yield a Rayleigh channel [20], which

makes a single-waveform component asymmetric, skewed, and non-Gaussian [20]. To deal with these issues, other fitting models are studied [19]. Later researchers found that the components can be neither Gaussian nor symmetric in measured FW signals, even when the emitted signal is Gaussian or symmetric. In [20], a nonparametric mixture model is studied to represent the waveform components. The author also applied a cluster analysis for echo decomposition. However, that method is time-consuming and requires manual fine-tuning steps.

In the second step, various filters such as Gaussian, low-pass, and Wiener filters have been applied to preprocess the received waveform signals in [16], [21], and [22]. Fine-tuning operations are needed for filter-based methods. But the denoising performance is not very convincing. Later, wavelet analysis [17], [23], [24] and empirical mode decomposition (EMD) [25] are used for this step. However, it is short of a general principle to select a wavelet basis function for wavelet analysis methods. Using EMD-based methods, signals are processed by removing high-frequency components for de-noising. But the selection of these noisy components heavily depends on specific operators.

For the optimization step, it is the final part of conventional decomposition methods. Optimization methods, such as LM [18], expectation maximization (EM) [26], and reversible jump Markov Chain Monte Carlo (RJMCMC) [27] are often used. However, in the initialization of these methods, the number of echo components in a received waveform signal needs to be prescribed. Additionally, conventional methods cannot decompose a large number of FW signals simultaneously because of additional iterative-based optimization steps. Furthermore, these optimization methods are time consuming and easy to be trapped into local optimal solutions.

With the rapid development of deep learning technology, many excellent AI models have been proposed to deal with different tasks including LiDAR signal processing. But before discussing the models for LiDAR signal, we review several SOTA works in RS data classification due to the similarity between the task and the first part of our method for FW signal decomposition. In [28], Wu et al. developed a novel convolution neural network with an advanced cross-channel reconstruction module, called CCR-Net, which can learn more compact fusion representations of different RS data. Compared to the SOTA multimodal RS data classification methods in 2020, CCR-Net is superior in achieving effective information exchange and compact fusion. In [29], Hong et al. proposed a general multimodel deep learning framework with a focus on RS image classification, which shows different fusion strategies as well as gives suggestions on how to train deep neural networks and build a neural network architecture. In [30], Hong et al. systematically investigate CNNs and GCNs in terms of hyperspectral image classification. Considering that traditional GCNs always suffer from a huge computational cost, the authors proposed miniGCNs that could be trained in minibatch fashion and allow for a straightforward inference of large-scale, out-of-samples datasets. In [31], Wu et al. proposed a multisource active fine-tuning vehicle detection (Ms-AFt) for vehicle detection of multisource RS data. Ms-AFt combines the transfer learning, active classification, and objects segmentation into a unified

framework. Also, Ms-AFt is a designed active classification network, which helps to get a progressive improvement in vehicle detection. In [32], Wu et al. provided a comprehensive survey of deep-learning-based object detection and tracking for unmanned aerial vehicle, which gives readers the instructive information about current state-of-the-art models.

Another task that we would discuss briefly is hyperspectral unmixing. Hyperspectral unmixing is a little similar to the FW LiDAR echo decomposition. Some hyperspectral unmixing methods assume that spectral variability follows a given probability distribution while the convention echo decomposition methods assume that the desired echo components follow the Gaussian or Gaussian-like distribution. Then, the echo decomposition problem can be translated into a function approximation task for the conventional decomposition methods; we call it model-driven approach. Hong et al. [33] propose an augmented linear mixing model (ALMM) for hyperspectral unmixing, where the scaling factors are modeled by the endmember dictionary and an addition dictionary. Also, it explores a data-driven dictionary learning method and gets satisfying performance compared to the previous state-of-the-art methods. In [34], Hong et al. proposed an endmember-guided unmixing network (EGU-Net) and developed a general deep learning framework to address the issue of nonlinear blind hyperspectral unmixing in a more effective and generalized way, thus avoided to generate the physically meaningless endmembers in practice.

With the development of deep learning technology, recently researchers apply neural networks for processing the FW LiDAR signals. In [35], Liu and Ke proposed a CNN-based superresolution network for FW LiDAR. In [36], Liao et al. proposed a deep-learning-based fusion framework to combine the complementary information from hyperspectral and FW LiDAR data for tree species mapping. In [37], Shinohara et al. tried to develop an end-to-end neural network model (Point2wave) for translating 3-D point clouds into their missing waveforms of FW LiDAR data using the SR-GAN. Also, the experimental results showed that Point2wave is able to translate the 3-D point clouds into desired waveform signals, and the translated waveform signals achieved nearly the same classification performance as the real waveforms. And in [38], Pashaei et al. proposed a DCNN-based classifier for waveform feature classification. The potential of raw samples from FW terrestrial laser scanning systems was explored for point cloud classification in city and countryside environments. In [14], Zorzi et al. design a neural network that uses FW signals to classify reflective targets, such as power lines, trees, buildings, etc. In [39], dense and residual neural networks are used to preprocess the received FW signals to generate nice preprocess results as mentioned before. Then, optimization methods are used for LiDAR signal decomposition. In [40], Aßmann et al. come up with the LiDARNet to do superresolution detection and fast peak localization using complicated multireturn waveform signals measured by the photon detector arrays. However, this work relies on the echo component assumption of the Gaussian model. It also ignores other important information of a waveform component, such as shape, except the center location.
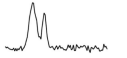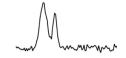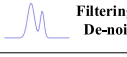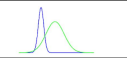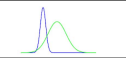
Fig. 1.　While conventional methods use multiple steps to decompose the waveform components, AFD-Net is an end-to-end neural network, which could get the decomposition result directly.

Fig. 2.　Many real examples show that the waveform components can be neither Gaussian nor symmetric, even the emitted signal is Gaussian.

Note that the above works by using deep networks are still initial for LiDAR signal decomposition. In this article, an end-to-end echo decomposition network, Attention Full-waveform Decomposition Network (AFD-Net), is designed. We do not assume echo components following any parametric models or functional distributions. A received waveform signal is sent into the network. The number of echo components is estimated first and then decomposed echo components are output. As shown in Fig. 1, no more extra step is needed in AFD-Net. To our best knowledge, we are the first to design an end-to-end network for the FW echo decomposition task. Compared to existing methods, our contributions can be summarized as follows.

1) End-to-end FW echo decomposition based on a self-attention-based neural network is proposed.
2) No assumption on waveform components such as parametric models or symmetric distributions is required in our decomposition approach.
3) No prior knowledge on the number of echo components is required in the decomposition methods. The number is estimated automatically through the classification module in AFD-Net.
4) We develop a tool, LiDAR-Gen, to generate the synthetic waveform signals, which can solve the problem of shortening datasets for the LiDAR waveform decomposition task.

The rest of this article is structured as follows. Section II introduces the principle for FW echo decomposition and the datasets we designed. The architecture of the AFD-Net is described in Section III. Section IV shows the experiment results. Finally, we draw the conclusions in Section V.

II. DECOMPOSITION PRINCIPLE AND DATASETS

A. Echo Decomposition Principle

For LiDAR, the received laser power P_r can be defined as [16]

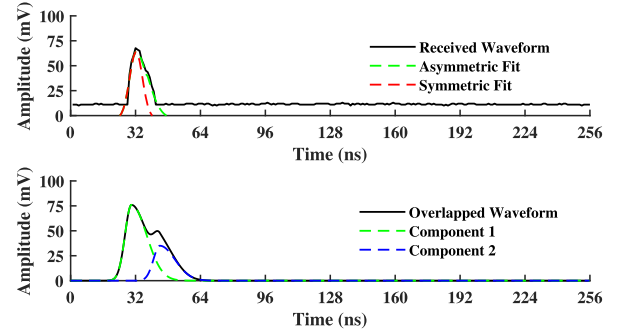$$P_r \propto \frac{\rho A_s}{\Omega L^4} P_t \qquad (1)$$

where P_t is the power of the output laser, ρ is the reflectivity of the scattering surface, A_s is the reflective area of a target, Ω is the divergence angle of a laser source beam, and L represents the transmission distance. It is clear that P_r is influenced by the four factors, L, A_s, ρ, and Ω. For FW LiDAR, we can get more detailed information about a reflective target [41]. The power $P_r(t)$ of the received waveform signals can be written as shown in

$$P_r(t) = \sum_{i=1}^{k} f_i(t) + \varepsilon(t) \qquad (2)$$

where $P_r(t)$ is still the power of a received waveform signal, and t is the sample time. $\varepsilon(t)$ in (2) represents the noise in a received waveform signal. If there are k reflective targets in the transmission path, then k echo components $f_i(t)$ can be found in $P_r(t)$. The goal of echo decomposition is to get the accurate $f_i(t)$ from a received waveform signal $P_r(t)$.

As mentioned before, the popular Gaussian model assumes that $f_i(t)$ follow the 1-D Gaussian distribution as shown in

$$f_i(t) = A_i e^{\dfrac{-(t-t_i)^2}{F_i^2/4\ln 2}} . \qquad (3)$$

A_i is the amplitude, t_i is the center location, and F_i is the FWHM of the ith Gaussian component. Then, the Gaussian-model-based methods try to extract accurate A_i, t_i, and F_i to generate echo components. However, the Gaussian assumption is too ideal to deal with some real waveform examples, as shown in Fig. 2. Once again, in our approach, we do not constrain the echo components $f_i(t)$ with any parametric distributions. Instead, we get $f_i(t)$ through the learning-based neural network AFD-Net, which is more suitable for the real-world waveform signals.

B. Datasets

Lack of suitable datasets is a big issue for many learning-based methods. FW echo decomposition faces the same problem since there is no public available datasets for the task. In this work, we propose a tool FW-GenTools to generate synthetic FW LiDAR signals with known peak number and realistic component shape. By doing so, it can save time for manually labeling received waveform signals, and avoid manual labeling errors.
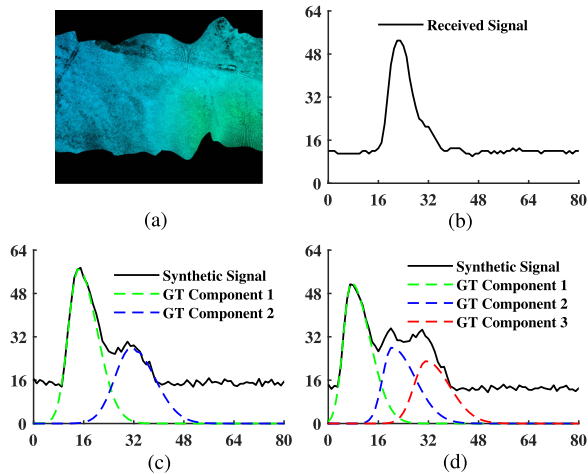
Fig. 3. (a) Scanned area in Yellowstone National Park, Wyoming. (b) Typical received FW signal in the scanned area. (c) Synthetic waveform signal consists of two symmetric waveform components (Ground-truth components 1 and 2). (d) Synthetic waveform signal consists of three asymmetric waveform components (Ground-truth components 1–3). The unit of the abscissa is nanoseconds, and the unit of the ordinate is millivolts.

We use the received waveform from NEON (i.e., The National Ecological Observatory Network, website [42]) as a reference to generate the synthetic FW signals for network training. Totally, 20 000 synthetic waveform signals are in the training set, in which 20% echo components are symmetric and 80% components are asymmetric. As shown in Fig. 3, panel (a) is the scanned area in Yellowstone National Park, Wyoming, and panel (b) shows a typical received waveform signal with a flat tail. Panels (c) and (d) show the synthetic signal in our training set, where (c) contains two symmetric waveform components and (d) contains three asymmetric waveform components. For the symmetric components, we use Gaussian and generalized Gaussian as models, because usually, the laser source pulse has a Gaussian waveform. For the asymmetric ones, received FW signals become asymmetric due to the intrinsic noise in electrical devices and the multiple transmission path effects [20]. The intrinsic noise introduces the flat tail into a waveform while the multiple transmission path effects lead the waveform components to be skewed and non-Gaussian. Indeed, the shape of most received waveform signals from NEON is asymmetric. We generate the asymmetric waveform signals by enlarge the width of the component gradually; thus, the flat tail appears in the synthetic waveform component. As for the number of echo components, few FW LiDAR signals have five or more echo components. Thus, in the training set, we set 10% waveform signals to have one echo component, 35% to have two components, 35% to have three components, and 20% to have four components.

For network training, the synthetic waveform signals and the corresponding ground-truth components are sent into AFD-Net to learn waveform features. After training, the received waveform signals from the real world are sent into the network to get the desired waveform components. More details about AFD-Net are introduced in the next section.

III. ARCHITECTURE OF AFD-NET

As shown in Fig. 4, the network AFD-Net consists of two modules, 1) the classification module and 2) the decomposition module. The classification module defines the number of echo components in a waveform signal, which helps the decomposition module output the correct number of desired waveform components.

A. Self-Attention Mechanism in 1-D Waveform Processing

In both the classification and the estimation modules, we use self-attention layers. Thus, we discuss the self-attention mechanism in the waveform signal processing first.

Self-attention is an attention mechanism, which calculates the representation of a sequence by encoding the positions of elements contained in the sequence. Since the self-attention based model is proposed by a team in Google in 2017, more and more models based on attention mechanisms have been developed recently. It has been used successfully in various tasks including natural language processing, machine translation, and image processing [43], [44].

For many tasks, the original input data sometimes are massive. It is hard for neural networks to encode the complete information into a single vector for training and learning. Self-attention matrix is defined to estimate the correlation between one element and other elements, which can help to encode the data sequence into a fixed length without using the whole original information. In the training step, self-attention-based models have the ability to learn information selectively from the features already extracted in the previous training epoch. Also, self-attention architecture can learn features from historical inputs, which help the models converge faster in training steps.

An FW LiDAR signal usually can be represented as a sequence of data points, such as 256 points for an 8-b A/D convert. Thus, 1-D convolution kernels are well suited for processing an FW LiDAR signal. Self-attention-based architectures can also be specially designed.

We define the waveform signal X as shown in

$$X = [x_1, x_2, \ldots, x_{256}] \tag{4}$$

where $[x_1, x_2, \ldots, x_{256}]$ is the intensity of the waveform signal at the sample time, where the sample ratio is 1 GHz, the sample interval is 1 ns. Our attention-based model does not blindly process all elements obtained from the previous layer but finds out the elements, which are associated with the specified tasks (classification or decomposition) for further processing. We place the self-attention layer (see Fig. 5) after the multiple feature extraction layers in the classification module to get a more precise classification result. Also, we adopt the attention blocks (see Fig. 8) to build the encoder–decoder architecture for decomposing the desired waveform components in the decomposition module.

B. Classification Module

As our final goal is to decompose all the echo components from the waveform signal, it is of great importance to know
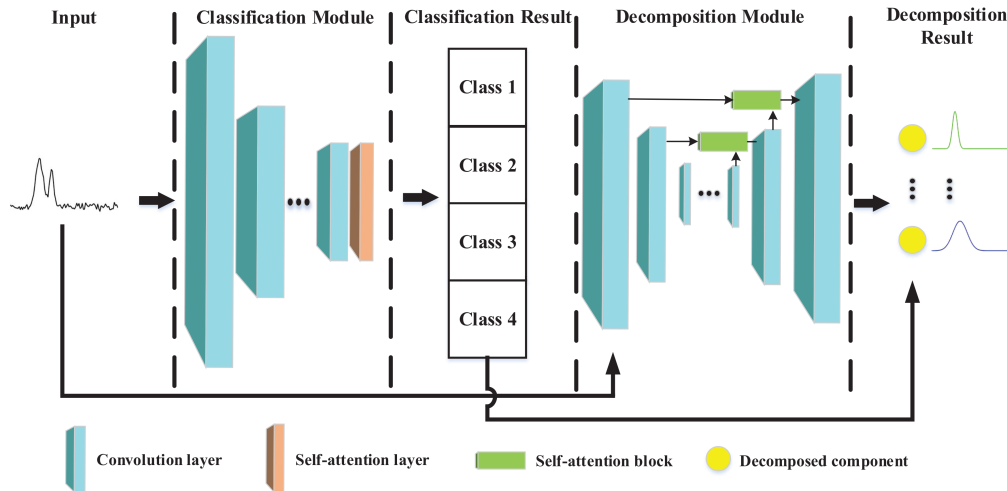
Fig. 4.   Architecture of the AFD-Net.
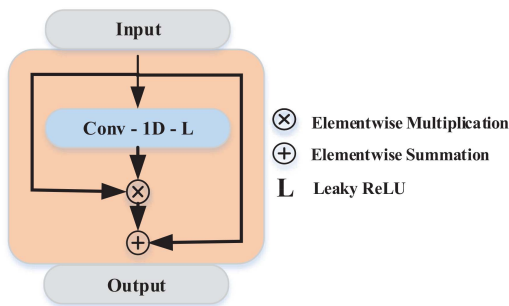

Fig. 5.   Self-attention layer in AFD-Nets classification module.
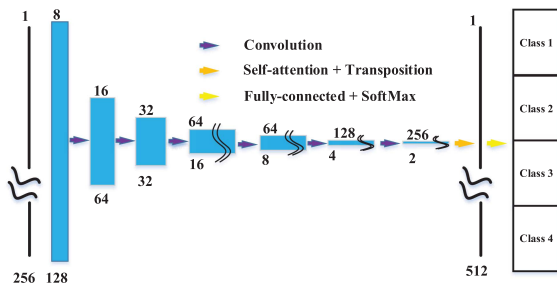

Fig. 6.   Architecture of the classification module, each blue box represents a feature map. The numbers above boxes correspond to a number of channels while numbers at the lower left edge correspond to the number of points in each channel.

the number of echo components contained in the waveform. Conventional methods usually estimate the number of the echo components by using additional complicated steps. But the estimation results are still limited. Some researchers also set the echo number manually, which always meets difficulty when dealing with massive data. In AFD-Net, we propose the CNN-based classification module to output the correct echo number. In our simulated experiments, the overall classification accuracy is 98.26%.

As shown in Fig. 6, the classification module consists of ordinary convolution layers, a self-attention layer, a fully-connected layer, and a SoftMax layer. In detail, the input waveform signals are sent into six convolution layers for feature extraction first. These convolution layers all have the same kernel size and padding parameter, where the kernel size is 5 and the padding stride is 2. Additionally, each convolution layer is followed by the ReLU activation and max-pooling, which reduce the data length by a factor of two. The self-attention layer is placed behind the multiple convolution layers as presented in Fig. 5. The extracted feature maps obtained from convolution layers are used to calculate an attention matrix, which helps the model to learn the correct echo number more efficiently. Finally, a SoftMax layer is applied for calculating the number of echo components contained in the waveform signal.

In the training step, we use the synthetic training dataset, which contains 20 000 waveform signals for training, as mentioned in Section II. Each waveform signal in the training set has the corresponding label that indicates the number of echo components in the waveform signal. Besides, the length of all the waveform signals is 256. The maximal intensity of the waveform signals is normalized before sending to the network. Thus, our model makes judgment based more on the waveform shape, rather than the intensity of received waveform signals. We apply the cross-entropy loss as the loss function in a classification module. Also, we use the Adam algorithm with a cyclic learning rate (cosine annealing) for optimization. Additionally, we find that adding a dropout layer before SoftMax could improve the final accuracy, when the dropout probability equals 0.5, the classification module gets the best performance.

C. Decomposition Module

As shown in Fig. 7, we apply the neural network with a modified U-Net architecture [44], [45] for decomposing the waveform components. U-Net-based neural networks have made great progress in the biomedical image segmentation tasks. Waveform decomposition is a little similar to image segmentation, which tries to assign a correct class label to each pixel. However, when waveform components overlap, each element in received
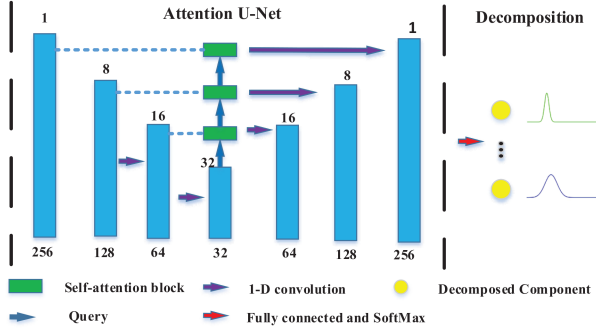
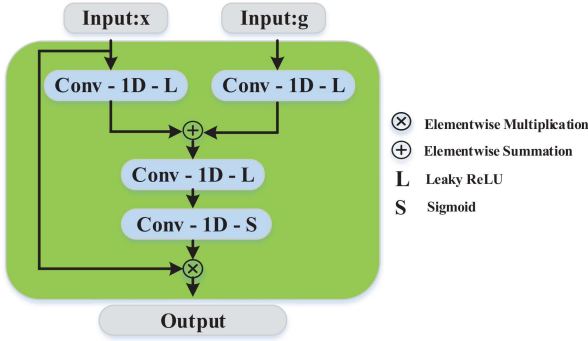Fig. 7. Architecture of the decomposition module in AFD-Net.

Fig. 8. Self-attention block in AFD-Nets estimation module.

waveform signals cannot be assigned with a unique class label. Traditional segmentation neural networks are not able to handle such a situation. We modified the architecture of U-Net to solve the overlapping issue. In our AFD-Net, the last layer does not output the segmentation map anymore. It is replaced by the decomposition layer that outputs the desired waveform components directly. The number of kernels in the last decomposition layer is set according to the result of the classification module. Each kernel outputs a single-waveform component.

Additionally, specially designed self-attention blocks for waveform processing are applied in the decomposition module, as shown in Fig. 8. Self-attention blocks calculate the attention coefficients, which lead the model to be more specific to local regions, and then improve the decomposition performance compared to the experimental results obtained without using self-attention blocks. Here, we denote the ith input feature map as x^i, and the gate information from the previous layer (see Fig. 7) as g^i. The multidimensional attention coefficients α^l are formulated as shown in the following equations [44]:

$$q_{\text{att}}^i = \sigma_1(W_x^T x^i + W_g^T g^i + b) \tag{5}$$

$$\alpha^l = \sigma_2(q_{\text{att}}^i(x^i, g^i, \Theta_{\text{att}})) \tag{6}$$

where $\sigma_2 = \frac{1}{1+\exp(-x^i)}$ is the sigmoid activation function. The attention block is characterized by a series of parameters Θ_{att}, which contains the linear transformation matrixes W_x and W_g, and the bias terms b. The linear transformations in self-attention blocks are computed by using the channelwise $1 \times 1 \times 1$ convolution kernels. Some research works design attention layers

that apply the SoftMax activation function for normalizing the attention coefficients. However, using SoftMax sequentially always leads to a sparse output. To avoid this, we choose sigmoid as the activation function in self-attention blocks.

We use the synthetic waveform signals with ground-truth waveform components to train the decomposition module. According to the classification result, the kernel number in the last layer is set automatically. Then, the waveform signals pass through the decomposition module, and the decomposed waveform components are output in the end. To characterize the decomposition errors, we use the L_2-norm distance between the decomposed waveform components and corresponding ground-truth as the loss function. As for optimization, Adam optimizer is applied in the training step of the decomposition module.

IV. EXPERIMENT RESULT

A. Evaluation Criteria

We evaluate AFD-Net using simulated and measured data. For the classification module, we define the classification accuracy as defined in

$$\text{Acc} = N_0/N \tag{7}$$

where N is the total number of the waveform signals sent for decomposition, and N_0 is the number of waveform signals classified with correct labels.

For the decomposition module, we use the determination coefficient R^2 to evaluate the performance of waveform decomposition as defined in

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left[P_r(t_i) - \sum_{j=1}^{M} \text{Comp}_j^{\text{dec}}(t_i) \right]^2}{\sum_{i=1}^{N} \left(P_r(t_i) - \overline{P_r} \right)^2}, \tag{8}$$

where N is the number of samples in the received waveform signal, $P_r(t_i)$ is the intensity of the ith sample in the received FW signal, $\overline{P_r}$ is the mean of the received waveform signal, $\text{Comp}_j^{\text{dec}}(t_i)$ is the intensity of the ith sample in the jth decomposed waveform component, and k is the number of the waveform components decomposed from the received waveform signal.

As shown in (2), noise $\varepsilon(t)$ exists in received waveform signals. We are able to get the exact ground-truth waveform components in the simulation experiment. So we replace $P_r(t_i)$ with multiple $\text{Comp}_j^{\text{GT}}(t_i)$ to calculate the determination coefficient R^2 as shown in the following equations, which leads to a more precise evaluation result in the simulation experiments:

$$P_r(t_i) = \sum_{j=1}^{k} \text{Comp}_j^{\text{GT}}(t_i) \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left[\sum_{j=1}^{k} \text{Comp}_j^{\text{GT}}(t_i) - \sum_{j=1}^{k} \text{Comp}_j^{\text{dec}}(t_i) \right]^2}{\sum_{i=1}^{N} \left[\sum_{j=1}^{k} \text{Comp}_j^{\text{GT}}(t_i) - \overline{\sum_{j=1}^{k} \text{Comp}_j^{\text{GT}}(t_i)} \right]^2}. \tag{10}$$

In general, the closer R^2 is to one, the decomposed waveform components are more similar to the ground-truth waveform
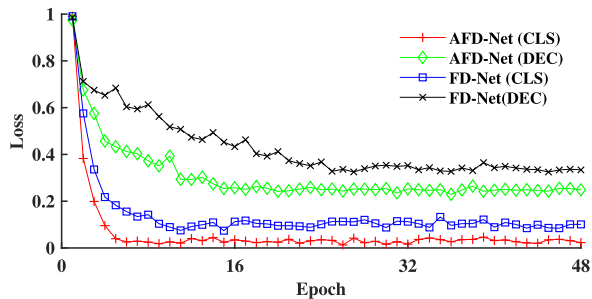
Fig. 9.   Training details of AFD-Net and FD-Net.

TABLE I
CLASSIFICATION RESULT USING SIMULATED DATA

Class	1	2	3	4	Overall
AFD-Net Acc (%)	99.40	98.51	98.23	97.30	98.26
FD-Net Acc (%)	98.20	97.89	97.43	96.20	97.40

components, thus the more accurate the decomposition result is [46].

B. Experiments With Simulated Data

1) Ablation: We do some ablation experiments based on the simulation data to prove that self-attention layer in the classification module and self-attention block in the decomposition module help our model work more efficiently. We define the neural network FD-Net (FW decomposition neural network) by removing the attention layers and blocks from AFD-Net. Then we train both FD-Net and AFD-Net using the same training dataset. The figure of the loss value versus epoch during the training process is shown in Fig. 9, where CLS means the classification module, and DEC means the decomposition module.

Obviously, the loss of AFD-Net drops faster than FD-Net, both in the classification module and decomposition module. Besides, AFD-Net takes less epochs to reach the state of convergence compared to FD-Net, which means the attention matrix calculated by attention layers and blocks help the model to complete the classification and decomposition tasks more efficiently. Ablation results and analysis of the designed attention modules are shown in the following sections.

For the number of parameters, totally AFD-Net has about 4.5 million parameters, where the classification module has about 0.5 million parameters and the decomposition module has 4 million. As for training details, it takes about 14 h for the training step using a single RTX-3090 GPU card.

2) Classification: In the simulation experiment, 5000 waveform signals with known component numbers are used for testing the classification accuracy of AFD-Net and FD-Net. As shown in Table I, the overall accuracy of AFD-Net in the simulation experiments is 98.26%. As the number of waveform components increases, the classification accuracy drops a little. But AFD-Nets classification accuracy is always higher than FD-Net. From Fig. 10, we can find that when the waveform signals contain one echo component, the classification accuracy of AFD-Net reaches the highest. Only very few signals are incorrectly classified with label 2. For the waveform signals
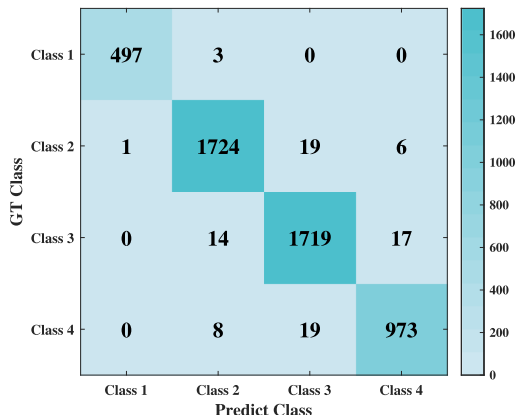

Fig. 10.   Confusion matrix of AFD-Nets classification result using simulated data.

TABLE II
DECOMPOSITION RESULT USING SIMULATED DATA

Class	1	2	3	4	Overall
AFD-Net (μ_{R^2})	0.9962	0.9941	0.9934	0.9881	0.9948
AFD-Net (σ_{R^2})	0.4716	0.8743	0.9316	1.3276	0.9447
FD-Net (μ_{R^2})	0.9957	0.9935	0.9927	0.9873	0.9922
FD-Net (σ_{R^2})	0.5373	0.9134	1.1067	1.2897	1.0187

containing two echo components, false outputs are more likely to assign the waveform signals with label 3. For the waveform signals containing three echo components, the wrong decisions always regard the signals containing two or four echo components. When it comes to class 4 or waveform signals containing four components, the task is more difficult. Thus, AFD-Net presents the lowest accuracy 97.30%, compared to other classes. However, 97.30% is a very good result for the classification task. As shown in Figs. 11 and 12, because we do not make any constrain to the shape of waveform components, for both symmetric and asymmetric components, AFD-net presents excellent classification performance. In addition, AFD-Net is able to output precise classification results when there are overlapped components in waveform signals, which is a hard occasion for convention decomposition methods to deal with.

3) Decomposition: To evaluate the decomposition performance of AFD-Net and FD-Net, 5000 synthetic waveform signals with known ground-truth components are used. We use the determination coefficient R^2 to compare the difference between the decomposed waveform components and the ground-truth components. The more closer R^2 is to 1, the better the decomposition result is. The mean μ_{R^2} and the standard deviation σ_{R^2} of R^2 are summarized in Table II. The distribution of R^2 is shown in Fig. 13. We can find that AFD-Net performs better than FD-Net all the time. Also, as shown in Figs. 11(a) and 12(a), when the waveform signal contains one echo component, the decomposition module in AFD-Net is able to extract the waveform component from noise efficiently. When it comes to the more complex occasions as shown in Figs. 11(b)–(d) and 12(b)–(d), AFD-Net decomposes the components from the waveform signals successfully even though, there are overlapping or the waveform components are not ideally symmetric.
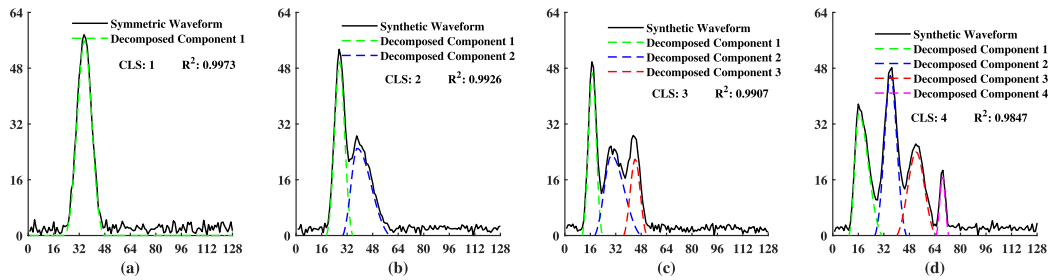
Fig. 11. Decomposition result of AFD-Net using simulated data. (a) Waveform signal contains a single symmetric component. (b) Waveform signal contains one symmetric component and one asymmetric component. (c) Waveform signal contains one symmetric component and two symmetric components. (d) Waveform signal contains two symmetric component and two asymmetric components. The unit of the abscissa is nanoseconds, and the unit of the ordinate is millivolts.
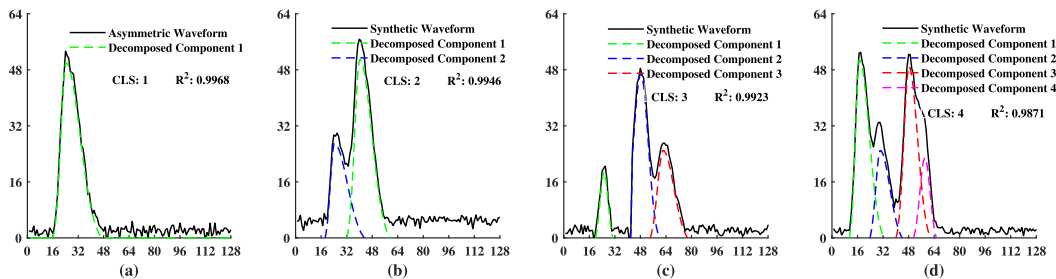

Fig. 12. Decomposition result of AFD-Net using simulated data. (a) Waveform signal contains single asymmetric component. (b) Waveform signal contains two asymmetric components. (c) Waveform signal contains two symmetric components. (d) Waveform signal contains one symmetric component and three asymmetric components. The unit of the abscissa is nanoseconds, and the unit of the ordinate is millivolts.
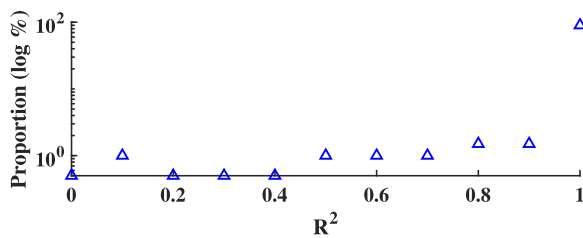

Fig. 13. Distribution of R^2 intervals for AFD-Net using simulated data.

TABLE III
CLASSIFICATION RESULT USING MEASURED DATA

Class	1	2	3	4	Overall
AFD-Net Acc (%)	100.00	95.71	95.71	95.30	98.26
FD-Net Acc (%)	95.00	92.86	91.42	92.5.0	92.50

Fig. 14. Confusion matrix of AFD-Nets classification result using measured data.

Fig. 15. Distribution of R^2 intervals for AFD-Net using measured data.

C. Experiments With Measured Data

The waveform signals discussed in this section are obtained using an Optech Gemini sensor as listed in NEON. The scanned area is in Yellowstone National Park in Wyoming. The LiDAR data are provided in a binary format defined by NEON. After changing the data format, the slant-range waveform is relative signal intensity versus nsec time bins.

1) Classification: To evaluate AFD-Nets classification performance on measured data, we manually annotated the components number of the 200 received waveform signals. The classification result is shown in Fig. 14 and Table III. For the waveform signals that contain one echo component, our AFD-Net makes all the correct classification predictions. As the
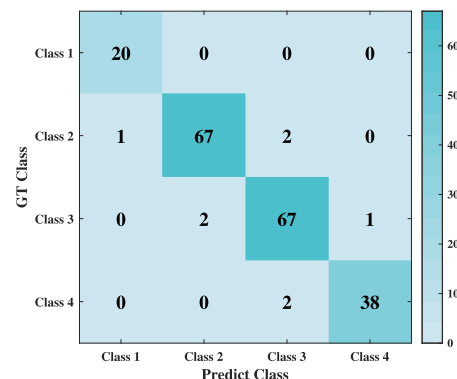
number of waveform components increases, although AFD-Net classification accuracy drops a little, it still has the overall accuracy of 96% for the measured waveform signals, which
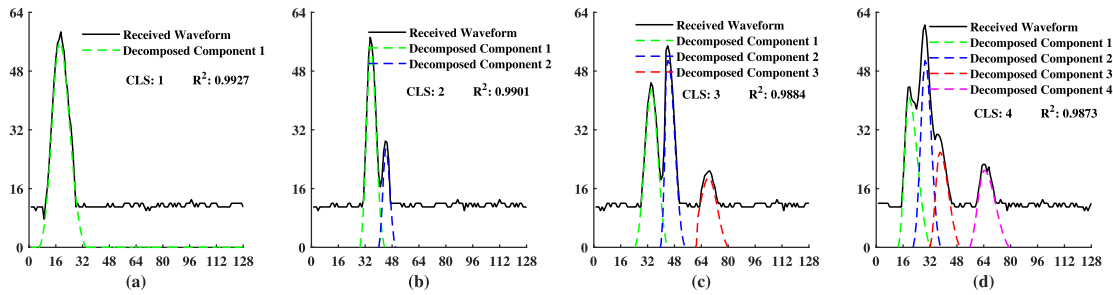
Fig. 16. Decomposition result of AFD-Net using measured data. (a) Received waveform contains one component. (b) Received waveform contains two components. (c) Received waveform contains three components. (d) Received waveform contains four components. The unit of the abscissa is nanoseconds, and the unit of the ordinate is millivolts.

TABLE IV
DECOMPOSITION RESULT USING MEASURED DATA

Class	1	2	3	4	Overall
AFD-Net (μ_{R^2})	0.9865	0.9847	0.9774	0.9726	0.9799
AFD-Net (σ_{R^2})	0.7516	0.9213	1.1031	1.1676	1.0179
FD-Net (μ_{R^2})	0.9841	0.9827	0.9746	0.9703	0.9775
FD-Net (σ_{R^2})	0.8614	1.0424	1.2174	1.4018	1.1574

is a very good classification result and can ensure the subsequent decomposition module to decompose waveform signals successfully.

2) Decomposition: Considering noise $\varepsilon(t)$ exists in the received waveform signals, it is predictable that the value of R^2 will drop as 8 is used to calculate the determination coefficient R^2. Table IV shows μ_{R^2} and σ_{R^2} of the decomposition results using AFD-Net and FD-Net with measured data. We find that the decomposition result of AFD-Net is also better than FD-Net. Self-attention blocks help the decomposition module to get better performance. Fig. 15 shows the distribution of R^2. It is obvious that the interval of 0.9–1 accounts for most of the cases. Fig. 16 shows the decomposition results of AFD-Net for the measured waveform signals. For some complicated occasions in which echo components overlap, AFD-Net is able to decompose the components from the received waveform signals.

V. CONCLUSION

In this article, we propose a self-attention-based neural network named AFD-Net for decomposing the waveform components from received FW LiDAR signals. Compared to conventional decomposition methods, our approach has the following advantages. AFD-Net is an end-to-end decomposition method while conventional methods need multiple steps to decompose the components from waveform signals. Additionally, we do not make any constrain to the distribution of the waveform. Thus, our decomposition method is more suitable for the real-world measured waveforms compared to the methods using functional fitting models. As for the decomposition efficiency, our deep-learning-based method can decompose serials of received waveform signals at the same time after the training step, it is much more efficient than iterative-based conventional methods. Besides, the specially designed self-attention architecture also speeds up the training process. Because there are no publicly

available datasets for the waveform decomposition task, we develop a tool named LiDAR-Gen to generate the synthetic waveform signals, which can help researchers to build their own datasets based on their specific scanning area.

For future directions, besides the network AFD-Net discussed in this article, graph convolutional networks (GCN) and transformers are theoretically feasible for the echo decomposition task. We wish that researchers would explore the potential of GCNs and transformers for the task in the future. Also, inspired by works such as in [34], we will explore new architectures of decomposition modules rather than current autoencoder-like ones. In addition, tasks such as canopy structure analysis, terrain slope estimation, shallow river bathymetry, etc., need high-accurate echo decomposition methods. AFD-Net has the potential to be used with practical application scenarios and then to get a better final result than conventional methods. Thus, another future direction is to design networks for specific application tasks.

REFERENCES

[1] H. Mitasova, M. F. Overton, J. J. Recalde, D. J. Bernstein, and C. W. Freeman, "Raster-based analysis of coastal terrain dynamics from multi-temporal LiDAR data," *J. Coastal Res.*, vol. 25, no. 2, pp. 507–514, 2009.

[2] P. Tarolli, "High-resolution topography for understanding earth surface processes: Opportunities and challenges," *Geomorphology*, vol. 216, pp. 295–312, 2014.

[3] C. Mallet and F. Bretar, "Full-waveform topographic LiDAR: State-of-the-art," *ISPRS J. Photogrammetry Remote Sens.*, vol. 64, no. 1, pp. 1–16, 2009.

[4] E. P. Baltsavias, "Airborne laser scanning: Basic relations and formulas," *ISPRS J. Photogrammetry Remote Sens.*, vol. 54, no. 2/3, pp. 199–214, 1999.

[5] X. Li, C. Liu, Z. Wang, X. Xie, D. Li, and L. Xu, "Airborne LiDAR: State-of-the-art of system design, technology and application," *Meas. Sci. Technol.*, vol. 32, no. 3, 2020, Art. no. 032002.

[6] G. Mandlburger, H. Lehner, and N. Pfeifer, "A comparison of single photon and full waveform LiDAR," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 4, pp. 397–404, 2019.

[7] S. Sun and C. Salvaggio, "Aerial 3D building detection and modeling from airborne LiDAR point clouds," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1440–1449, Jun. 2013.

[8] H. V. Duong, M. A. Lefsky, T. Ramond, and C. Weimer, "The electronically steerable flash LiDAR: A full waveform scanning system for topographic and ecosystem structure applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4809–4820, Nov. 2012.

[9] F. M. Danson et al., "Developing a dual-wavelength full-waveform terrestrial laser scanner to characterize forest canopy structure," *Agricultural Forest Meteorol.*, vol. 198, pp. 7–14, 2014.

[10] H. Zhou et al., "Estimation of canopy height using an airborne Ku-band frequency-modulated continuous waveform profiling radar," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3590–3597, Oct. 2018.

[11] R. Guliaev, V. Cazcarra-Bes, M. Pardini, and K. Papathanassiou, "Forest height estimation by means of tanDEM-X inSAR and waveform LiDAR data," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3084–3094, Feb. 2021.

[12] A. Chauve et al., "Advanced full-waveform LiDAR data echo detection: Assessing quality of derived terrain and tree height models in an Alpine Coniferous Forest," *Int. J. Remote Sens.*, vol. 30, no. 19, pp. 5211–5228, 2009.

[13] C. Mallet, F. Bretar, M. Roux, U. Soergel, and C. Heipke, "Relevance assessment of full-waveform LiDAR data for urban area classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 66, no. 6, pp. S71–S84, 2011.

[14] S. Zorzi, E. Maset, A. Fusiello, and F. Crosilla, "Full-waveform airborne LiDAR data classification using convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8255–8261, Oct. 2019.

[15] J. Jung, E. Pasolli, S. Prasad, J. C. Tilton, and M. M. Crawford, "A framework for land cover classification using discrete return LiDAR data: Adopting pseudo-waveform and hierarchical segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 2, pp. 491–502, Feb. 2014.

[16] M. A. Hofton, J.-B. Minster, and J. B. Blair, "Decomposition of laser altimeter waveforms," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 4, pp. 1989–1996, Jul. 2000.

[17] D. Li, L. Xu, and X. Li, "Full-waveform LiDAR echo decomposition based on wavelet decomposition and particle swarm optimization," *Meas. Sci. Technol.*, vol. 28, no. 4, 2017, Art. no. 045205.

[18] J. J. Moré, "The Levenberg-Marquardt algorithm: Implementation and theory," in *Numerical Analysis*. Berlin, Germany: Springer, 1978, pp. 105–116.

[19] A. Chauve, C. Mallet, F. Bretar, S. Durrieu, M. Pierrot-Deseilligny, and W. Puech, "Processing full-waveform LiDAR data: Modelling raw signals," in *Proc. ISPRS Workshop Laser Scanning SilviLaser*, 2007, vol. 36, pp. 102–107.

[20] Q. Li, S. Ural, J. Anderson, and J. Shan, "A fuzzy mean-shift approach to LiDAR waveform decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7112–7121, Dec. 2016.

[21] C. Wang, Q. Li, Y. Liu, G. Wu, P. Liu, and X. Ding, "A comparison of waveform processing algorithms for single-wavelength LiDAR bathymetry," *ISPRS J. Photogrammetry Remote Sens.*, vol. 101, pp. 22–35, 2015.

[22] Y. Qin, T. T. Vu, and Y. Ban, "Toward an optimal algorithm for LiDAR waveform decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 482–486, May 2012.

[23] C. Wang, F. Tang, L. Li, G. Li, F. Cheng, and X. Xi, "Wavelet analysis for ICESat/GLAS waveform decomposition and its application in average tree height estimation," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 115–119, Jan. 2013.

[24] C.-K. Wang et al., "Exploring weak and overlapped returns of a LiDAR waveform with a wavelet-based echo detector," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 39, pp. 1–6, 2012.

[25] Z. Zhang et al., "A novel noise reduction method for space-borne full waveforms based on empirical mode decomposition," *Optik*, vol. 202, 2020, Art. no. 163581.

[26] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.: Ser. B. (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[27] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[28] X. Wu, D. Hong, and J. Chanussot, "Convolutional neural networks for multimodal remote sensing data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Nov. 2021, Art. no. 5517010.

[29] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[30] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[31] X. Wu, W. Li, D. Hong, J. Tian, R. Tao, and Q. Du, "Vehicle detection of multi-source remote sensing data using active fine-tuning network," *ISPRS J. Photogrammetry Remote Sens.*, vol. 167, pp. 39–53, 2020.

[32] X. Wu, W. Li, D. Hong, R. Tao, and Q. Du, "Deep learning for unmanned aerial vehicle-based object detection and tracking: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 91–124, Mar. 2022.

[33] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[34] D. Hong et al., "Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2021.3082289.

[35] G. Liu and J. Ke, "Deep-learning for super-resolution full-waveform LiDAR," *Proc. SPIE*, vol. 11187, 2019, Art. no. 1118714.

[36] W. Liao, F. Van Coillie, L. Gao, L. Li, B. Zhang, and J. Chanussot, "Deep learning for fusion of apex hyperspectral and full-waveform LiDAR remote sensing data for tree species mapping," *IEEE Access*, vol. 6, pp. 68716–68729, 2018.

[37] T. Shinohara, H. Xiu, and M. Matsuoka, "Point2Wave: 3-D point cloud to waveform translation using a conditional generative adversarial network with dual discriminators," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11630–11642, 2021.

[38] M. Pashaei, M. J. Starek, C. L. Glennie, and J. Berryhill, "Terrestrial LiDAR data classification based on raw waveform samples versus online waveform attributes," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5702319.

[39] G. Liu and J. Ke, "Full-waveform LiDAR echo decomposition based on dense and residual neural networks," *Appl. Opt.*, vol. 61, no. 9, pp. F15–F24, 2022.

[40] A. Aßmann, B. Stewart, and A. M. Wallace, "Deep learning for LiDAR waveforms with multiple returns," in *Proc. IEEE 28th Eur. Signal Process. Conf.*, 2021, pp. 1571–1575.

[41] B.-C. Du et al., "A photon-counting full-waveform LiDAR," *Chin. Phys. Lett.*, vol. 36, no. 9, 2019, Art. no. 094201.

[42] National Ecological Observatory Network (NEON), "LiDAR slant range waveform (dp1.30001.001)," 2021. [Online]. Available: https://data.neonscience.org/data-products/DP1.30001.001/RELEASE-2021

[43] A. Vaswani et al., "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 1706–1717, 2017.

[44] O. Oktay et al., "Attention u-net: Learning where to look for the pancreas," in *Med. Imag. Deep Learn.*, 2018. [Online]. Available: https://openreview.net/forum?id=Skft7cijM

[45] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.

[46] L. Xu, D. Li, and X. Li, "A high success rate full-waveform LiDAR echo decomposition method," *Meas. Sci. Technol.*, vol. 27, no. 1, 2015, Art. no. 15205.

Gangping Liu (Member, IEEE) received the B.S. degree in optical engineering in 2018 from the Beijing Institute of Technology, Beijing, China, where he has been working toward the Ph.D. degree in optical engineering with the School of Optics and Photonics since September 2018.

His current research interests include full-waveform LiDAR signal processing and single-image superresolution.

Jun Ke (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the University of Electronic and Science Technology of China, Chengdu, China, in 1996 and 1999, respectively, the M.S. degree in mathematics from Purdue University, West Lafayette, IN, USA, in 2002, and the Ph.D. degree in optical and computer science from the Department of Electrical and Computer Engineering (ECE), University of Arizona (UA), Tucson, AZ, USA, in 2010.

She is currently the Associate Professor with the School of Optics and Photonics, Beijing Institute of Technology, Beijing, China. Her research interests include optical science and computational imaging.

Dr. Ke is a Senior Member of OSA and a Member of the Society of Photo-Optical Instrumentation Engineers (SPIE). She is the Reviewer of the *Journal of the Optical Society of America, Applied Optics, Optics Letters, Optics Express*, etc.