ADVANCED INTELLIGENT SYSTEMS
Open Access
www.advintellsyst.com

Generating Ultrasonic Foliage Echoes with Variational Autoencoders

Michael Goldsworthy and Rolf Müller*

Navigation through dense foliage presents a fundamental challenge to autonomous systems, and achieving a performance level similar to echolocating bats could have important applications in areas such as forestry and farming. However, the clutter echoes originating from such environments have been difficult to analyze. To study the problem of sonar-based navigation in dense foliage in simulation, an artificial generation system for leaf impulse responses (IRs) based on variational auto-encoders is proposed. The system is to aid the construction of artificial foliage echo environments. A dataset of leaf echoes was collected in an anechoic chamber and convolved with the original signal to estimate the IR of each leaf. A modified version of the conditional variational autoencoder - generative adversarial network (cVAE-GAN) architecture was trained successfully on this dataset to produce a generative model that was conditional on leaf viewing angles, size, and species. The IRs generated by the model capture quantitative and qualitative similarity to the measured IRs. It surpasses the previous state of the art foliage echo model based on reflecting disks. The model's computational efficiency and its success suggest its potential use for simulating large environments of foliage to study bat biosonar and aid in engineering biomimetic sonar devices.

1. Introduction

Autonomous navigation through complex natural environments such as dense foliage is a challenging task that could have a transformative impact if tackled successfully. The ability to navigate through foliage, for example, could enable applications in areas such as autonomous agriculture and forestry.[1] GPS-based navigation has been successfully tried in above-canopy drone navigation where it has been able to deliver a high degree of localization accuracy.[2] However, due to the obscuring effects of dense foliage, it is difficult to gather substantial amounts of data on what lies below the canopy based on images that are taken from above.[2] Hence, for tasks which require high precision observation of features located within or under a forest canopy, flights through the foliage under the canopy are likely necessary.

Any successful attempt to navigate in large-scale, complex, and GPS-denied environments would have to overcome four main obstacles[3]: 1) Operating on a large-scale makes generating maps (as needed by simultaneous localization and mapping (SLAM) based systems) infeasible, or at least very difficult, 2) the complexity makes planning difficult, 3) if the environment is dynamic, then pre-planning may be impossible, 4) sensors have limited sensing capacity, detecting all relevant information for navigation may be beyond what any given sensor is capable of, so encountering complex cluttered environments may confuse the unmanned aerial vehicle (UAV). Due to all these complexities, navigating through environments with cluttered natural foliage, environments which are large, complex, and dynamic, is a fundamental challenge for autonomy.

Many species of bats are capable of overcoming the challenges associated with navigation in complex natural environments based on sensory information provided by their biosonar systems. These species are able to hunt prey, avoid collisions, and return home almost entirely relying on echoes from the foliage in which they live.[4,5] For engineering, sonar as a sensing modality has several advantages over vision, such as operation in darkness and fog, as well as lower data rates and power requirements. However, foliage echoes are difficult to understand since they are typically composed of contributions from many – potentially hundreds to thousands – unresolved scatterers, i.e., leaves and other reflecting facets in a foliage, that each come in various sizes and shapes as well as present themselves to the sonar in different orientations. Because of a lack of knowledge regarding the components of these so-called clutter echoes, performing estimation tasks of detection, localization, and classification present substantial computational problems.[6]

Because of the persistent difficulties in interpreting clutter echoes, currently existing man-made sonar systems have not

M. Goldsworthy
Department of Computer Science
Virginia Tech
Blacksburg, VA 24060, USA
E-mail: michaeljg@vt.edu

R. Müller
Department of Mechanical Engineering
Virginia Tech
Blacksburg, VA 24060, USA

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aisy.202300697.

DOI: 10.1002/aisy.202300697

been able to achieve bat-like performance for sensing tasks in complex, natural environments. Technical sonar is designed to minimize beamwidth and hence increase its spatial resolution, which requires a large ratio of emission/reception aperture size to wavelength.[7] This is typically realized by virtue of an array of a large number of emitting and receiving elements. In contrast to this, bats only require three elements (their nose or mouth for emission, and two ears for reception). As a consequence of the animals' small size, the beamwidths of bats are substantially wider than those of engineered systems, suggesting the involvement of completely different sensing paradigms.[8] Due to their resolution-based approach, current technical sonar systems need to be much larger and require more computational power than bats. A better understanding of dense foliage echoes and the bat's biosonar sensing strategies for interpreting them could hence help advance sonar engineering techniques to yield more capable, as well as more parsimonious systems.

Additionally, acoustic/ultrasonic clutter signals appear in domains other than navigation in foliage, such as in shallow-water sonar sensing[9] and in medical ultrasound.[10] Hence, insights into how bat biosonar can take advantage of cluttered acoustic signals to achieve their goals could motivate the development of similar methods that could have a transformative impact on sensing technology operating in these domains.

It has been shown that echoes from foliage consist primarily of contributions from leaves rather than branches.[11] Plant leaves come in an enormous variety of shapes and sizes, and many natural foliage environments contain multiple constitutive plant species. Simulating the echoes from foliage realistically therefore requires a model with an amount of variation in the leaf types and arrangements that comes sufficiently close to real foliage. Physical simulations, such as by finite or boundary element methods, would require a mesh to represent each leaf with its given shape and size, necessitating the creation of hundreds or thousands of realistic and varied meshes, and requiring a substantial computational effort for each mesh and each acoustic viewing angle. Thus, any attempt to study bat biosonar by a simulation of acoustic reflections using a physics-based model in a cluttered natural environment would be hampered by the time and computational complexity of such a simulation. This motivates a need for a computationally feasible simulation environment that still retrains a sufficient degree of quantitative similarity to real foliage environments.

Our goal has thus been to create a simulation environment capable of generating realistic and varied impulse responses of entire trees which can be used to investigate the question of which signal parameters would be useful for biosonar and biomimetic sonar and perhaps discover alternative approaches to methods already established in fields such as man-made sonar or biomedical ultrasound. Our approach follows prior work by simulating impulse responses from individual leaves in a way that is computationally feasible and enables them to be added together to simulate entire tree echoes.

Leaf echoes have been previously simulated using deterministic methods based on idealized targets, such as point scatterers[12] and disks.[13] Clearly, point scatterers lack all of the features that are due to leaf properties other than location, in particular size, shape, and viewing angle. Tree echoes depend substantially upon these leaf properties, for example, the statistical properties of fig tree echoes versus yew tree echoes have been found to differ presumable due to the specular nature of the much larger, planar leaves of the fig tree.[6] Modelling leaves as disks includes size and viewing angle dependencies, but the entire diversity of shapes found in real leaves cannot be represented. Both of these methods ignore multiple bounces between leaves (Born approximation, a commonly used approximation in ultrasound[14,15]) and ignore the shadowing of deeper layers of leaves from shallower layers. To address the challenges of modelling foliage echoes in a more realistic fashion while keeping computational cost low, we propose a machine-learning approach to generate individual leaf impulse responses. To this end, we have assessed two recent generative deep learning methods: Generative adversarial networks (GANs[16]) and variational autoencoders (VAEs[17]) (**Figure 1**).

The key innovation of GANs is to combine two neural networks, one a generator and the other a discriminator, which are trained alternatively and in a competitive way.[16] The aim is to train the generator to produce simulated samples that are similar enough to the samples in a set of real data that the discriminator fails to distinguish between real and simulated samples. The generator takes a vector from a random distribution (typically a Gaussian) as input and outputs vectors that have the same shape as the target data. The discriminator then takes both the simulated and real samples as inputs, and is trained to distinguish between simulated and real samples. The generator is trained to make the discriminator guess wrong, thus it is trained

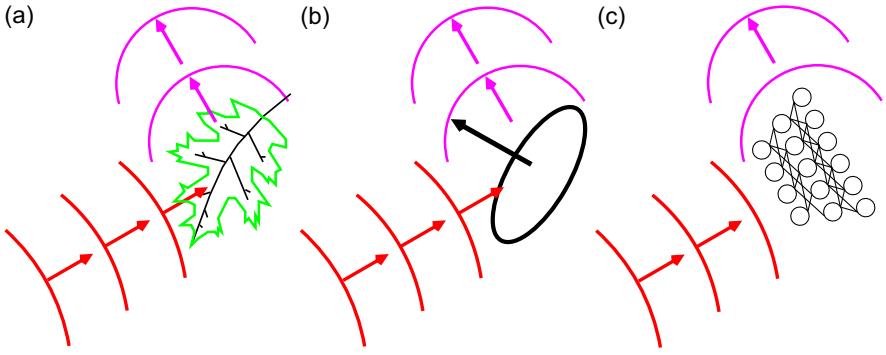

Figure 1. Schematic representations of the biological paragon and different echo modelling approaches: a) Example of the biological model: An oak leaf reflecting sound, b) disk model as a simple approximation of a leaf's acoustic response, and c) VAE model for creating artificial leaf responses.

to generate samples that are similar to the real elements in the dataset. GANs have undergone many improvements and an enormous number of variants exist as a result.[18,19] They have most prominently seen application in the generation of images and other image-related tasks, such as in-painting, super-resolution, and image-to-image translation.[19] Due to the periodic nature of audio, naive application of image generation methods to generating raw audio waveforms typically fails. GANs frequently fail to capture global dependencies between distant parts of images, and for periodic signals like soundwaves there is too high global correlation between all parts of the signal for naive GAN architectures to reliably generate natural soundwaves, and for longer signals such as music there are dependencies within stuctures at multiple timescales,[20] however alternative GAN architectures have been devised to deal with this, such as ref. [21].

VAEs also use two neural networks but they are arranged like an autoencoder,[17] where one network encodes samples into a latent space and the other samples from the latent space and reconstructs the original samples, both networks trained to reduce the reconstruction error. The innovation of VAEs is to shape the latent space towards a prior distribution (typically a Gaussian) and thus new unseen samples can be drawn from the latent space and "decoded" into new samples that did not exist in the original dataset. Like GANs, VAEs have seen a rich diversity of recent developments and applications. They are typically used for image tasks, but also have been used for protein design,[22] language models,[22] source separation,[23] finance,[23] and many others.

Both methods come with common well known shortcomings and failure modes: GANs frequently fail to capture the full variety of the real data and focus on generating samples that are similar to only a small subset of the data, a failure mode known as mode collapse.[18] VAEs are less likely to suffer from this problem, but are known to produce samples that are blurrier than the sharper images produced by GANs.[22] Another problem known as entanglement[24] arises when training conditional models, where generators or decoders learn to ignore class labels and let the latent space contain all information, entangling the known labels with the general unknown generative factors.

In order to ameliorate the shortcomings of both methods, the current work has employed a deep-learning generative method to produce simulated leaf impulse responses that closely follows the conditional variational autoencoder - generative adversarial network (cVAE-GAN,[25]). This method combines VAEs and GANs in order to overcomes some the problems associated with each of the combined methods. The cVAE-GAN has been trained with a dataset of leaf impulse responses (IR) samples that were collected with a sonar head that mimicked the basic function of bat biosonar. The output of the model has been evaluated quantitatively against statistical features of the measured leaf impulse responses.

2. Experimental Section

2.1. Data Acquisition

In order to gather the large experimental dataset that is needed for training a deep-learning model, single leaves were placed in an anechoic chamber (4.5 × 2.2 × 2.5 m) and suspended in open space by two parallel thin fishing lines to minimize any acoustic reflections not originating at the leaf (**Figure 2**).

The ultrasonic emitter-receiver unit was placed at a distance of 1 m from the leaf. The emitted pulse consisted of linear-frequency modulated carrier that was swept down from 105 to 5 kHz over a duration of 2 ms. The pulse was converted to an analog signal with a conversion frequency of 1.6 MHz and a resolution of 12 bits (Arduino DUE, Arduino SA, Chiasso, Switzerland). The analog pulse was then emitted from an electrostatic ultrasonic transducer (Series 600, SensComp Inc., Livonia, USA). The echo from the leaf was recorded by an

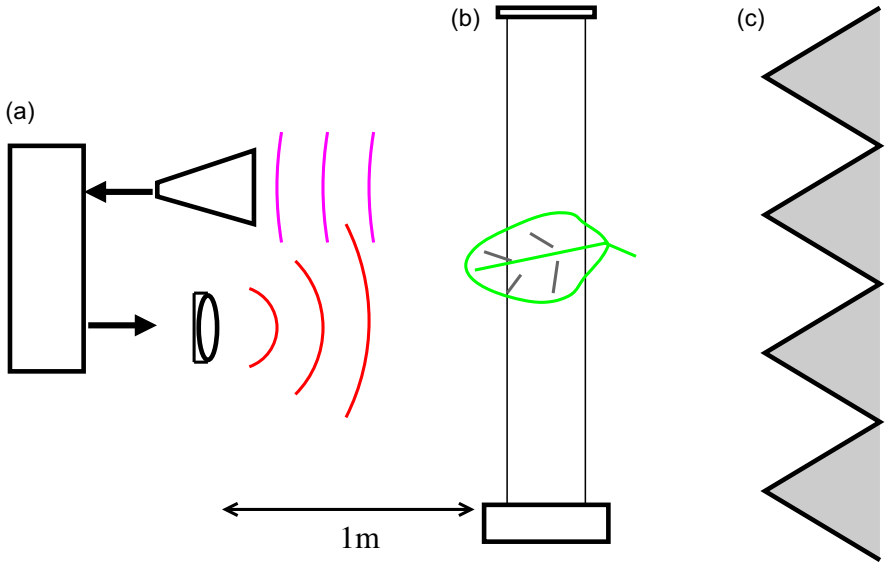

Figure 2. Experimental setup for measuring leaf impulse responses: a) ultrasonic emitter-receiver unit, b) leaf suspended with parallel fishing lines and rotated by a stepper motor, and c) anechoic chamber wall.

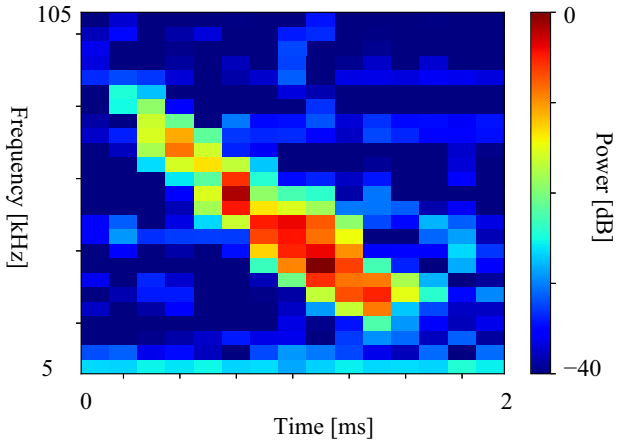

Figure 3. An example of a time-frequency plot of a chirp echo from a leaf before preprocessing.

ultrasonic microphone (Momimic, Dodotronic, Castel Gandolfo, Italy) over a recording duration of 25 ms, and then digitized with a sampling rate of 400 kHz and a resolution of 12 bits (Arduino DUE, Arduino SA, Chiasso, Switzerland). While recording each echo, the respective leaf's azimuth angle, elevation angle, size, and species were noted to generate the label set for the data (**Figure 3**).

The recorded echoes were subjected to matched filtering, i.e., convolved with the time-reverse of the digital chirp template to obtain an estimate of the leaf's impulse response, and clipped to a duration of 1 ms (i.e., 400 samples) centered on the maximum amplitude of the estimated impulse response. Amplitudes were normalized to fall in the range from −1 to 1 based on the minimum and maximum values in the overall data set. Similarly, all continuous labels (azimuth angle, elevation angle, size) were normalized into the range of 0 to 1. Leaf impulses that were in the bottom 10^{th} percentile of energy were discarded.

2.2. Generative Modelling

Since the chosen VAE architecture (i.e., cVAE-GAN) needed to be conditional, i.e., it needed the ability to generate impulse responses that depend on plant species and viewing direction, it was necessary to avoid entanglement in the latent space. Entanglement means that the latent dimensions are not independent of the factors that should control the generated outputs. This is especially a problem for conditional models where the conditional factors are at risk of getting subsumed by the latent space. For example, a possible case of entangling in the present work might be that the generational information for the azimuth angle is encoded by variations in the latent dimensions and hence the conditional information on azimuth given to the decoder network is ignored (i.e., is given zero weight in the network). To deal with entanglement, the chosen architecture was based on the disentangling version of the cVAE-GAN[26] that can be found in ref. [25]. In this architecture, an adversarial classifier was applied to the latent space and trained to classify the latent-space representations according to leaf species. Similarly, an adversarial regressor was trained to predict the other, continuous sample

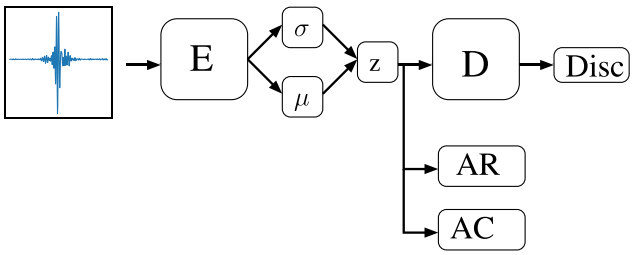

Figure 4. Overview of the components in the cVAE-GAN-based network used to generate the leaf impulse responses: E: encoder network, σ and μ: standard deviation and mean of the normal distribution from which the latent vector z is drawn, D: decoder network, AR: auxiliary regressor, AC: auxiliary classifier, Disc: discriminator. All component networks were used in training. For generation, only the decoder network was used.

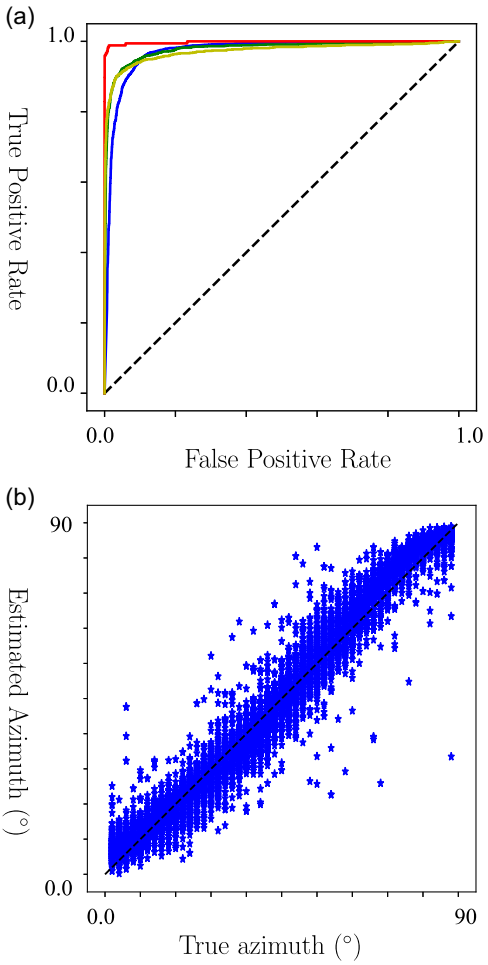

Figure 5. Accuracy of the neural networks used for classifying leaf species and estimating the azimuth angle of the training data: a) ROC curves of a neural network classifier trained to predict leaf species from the measured impulse responses, the four curves each correspond to one of the four species. b) Azimuth angles of measured IRs estimated by a regression network plotted against the respective true values.

labels (azimuth, elevation, and leaf size). The encoder contained a corresponding regularizing loss term which was opposed to the classifier and regressor, meaning that it was trained to make the

latent space unclassifiable. Instead of using a second latent code for the conditional information, following ref. [25], where the conditional information is mapped to its own latent space and trained to follow a multivariate Gaussian distribution, the conditional information was simply passed to the decoder in its original form (as in ref. [26]). The presence of these auxiliary networks was meant to ensure that the latent space did not contain this information and that the conditional information was passed to the decoder.

In the cVAE used here for the task of impulse response generation, the primary components were 1) an encoder that was trained to map signals to a Gaussian latent space and 2) a decoder to reconstruct the original signal. In addition, the cVAE contained two other networks designed to shape the latent space to avoid entangling: the auxiliary classifier and the regression network as described above. Since VAEs are known to produce blurrier images than GANs, there was an additional discriminator after the decoder to aid in the sharpness of the generated impulse responses by distinguishing between measured and generated samples (**Figure 4**). The decoder loss includes another term, which seeks to minimize the final discriminator's classification accuracy between generated and measured samples. The reconstruction error component of the loss function includes a weighting factor (γ), following,[27] which balances the reconstruction error with the regularization of the latent space. The specifics of the cVAE-GAN network architecture used were as follows: 1) Encoder: multilayer perceptron (MLP) with 3 hidden layers containing 300, 200, 100 nodes, respectively. 2) Decoder: MLP with 4 hidden layers containing 100, 200, 300, 400 nodes, respectively. 3) Discriminator: MLP with 2 hidden layers containing 200 and 10 nodes, respectively. 4) Auxilliary classifier and regressor: MLP with 2 hidden layers containing 100 and 5 nodes, respectively. 5) Activation functions: All hidden unit activation functions were rectified linear unit (ReLU), activation functions on the final layer we either sigmoid or softmax where appropriate. 6) Loss functions: Followed the method of ref. [25], with the addition of the a weighting factor (γ) applied to the reconstruction loss. 7) Optimizer: Adam.[28]

2.3. Analytical Metrics

For generation of impulse responses that are conditioned on the labels to be successful, the conditional information on the respective target property must be present in the samples from the experimental recordings. To establish whether this was the case, a deep-learning classifier was trained to determine the leaf species for the experimental samples, Similarly, a regression network was trained to predict the azimuth angle from the experimental estimates of the impulse responses. Success of these classification or regression experiments would establish that the experimental recordings contain the information that is necessary for a training a conditional VAE. Failure could mean that the information is either not contained in the data or the networks were not able to utilize it.

A standard one-dimensional convolutional neural network (CNN) classifier (6 layers of convolution with batch normalization, followed by 3 dense layers with dropout) was used on the training data to determine whether it can be classified by leaf species. Likewise, a straightforward MLP regression network (3 hidden layers consisting of 400, 20, and 10 neurons, respectively) was used to perform regression on the training data to predict the azimuth angle. After determining whether conditional information is accessible in the experimental recordings, the cVAE generator was trained to create synthetic impulse response data.

Based on this synthetic data, a number of metrics were used to analyze the generated samples and measure the performance of

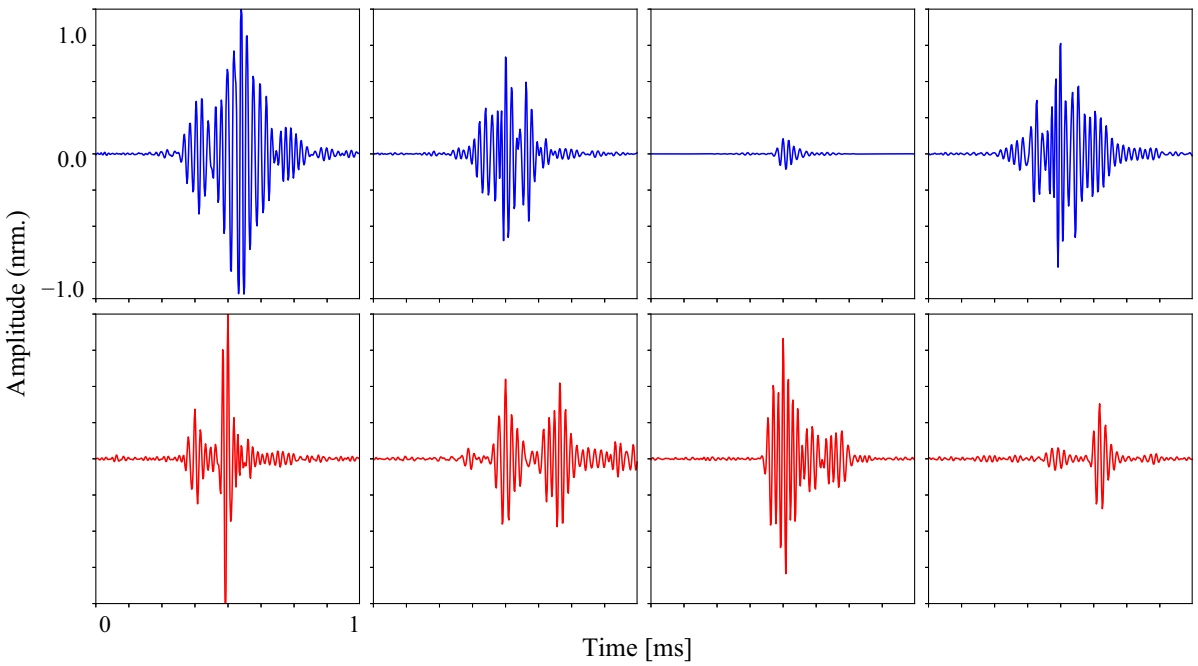

Figure 6. Examples of the waveforms for measured (top row) and cVAE-GAN-generated leaf impulse responses (bottom row).

the generative method, both qualitatively and quantitatively: As a first step, the generated IRs were visually inspected and compared qualitatively to the real IRs. Major failures of the generative method (e.g., mode collapse, excessive noise) could be detected qualitatively in this inspection.

Next, the variation in the signal energy (estimated as a sum of squares of the IR amplitudes) as a function of azimuth angle was compared across measured and generated IRs. Comparing the relationship between energy and azimuth angle was meant to ensure that meaningful and realistic conditioning is happening on the input azimuth angle label to the generator.

Furthermore, the cVAE generator was compared to the previous state of the art method of conditional leaf IR generation that used a disk model for the leaves.[13] To this end, the first three standardized moments (variance, skewness, and kurtosis) of the amplitude distributions of the IRs generated by the cVAE were compared to those associated with those of real IRs and simulated IRs generated with the disk model.

Finally, a regression network was trained to predict the azimuth angle for the real IRs in the training data, and was then used to predict the azimuth angle of the conditionally generated IRs. This experiment was intended to measures whether the azimuth-angle information contained in the signals is being meaningfully imparted on the synthetic impulses responses through our conditional generation approach. This process was also reversed, i.e., a regression network was trained on the generated IRs and then used to predict the azimuth angle of the measured data.

3. Results

The regressor for azimuth angle and the classifier for leaf species were successfully trained on the measured data in the training dataset. Both networks were found to perform with a high degrees of accuracy (**Figure 5**): The azimuth-angle regressor

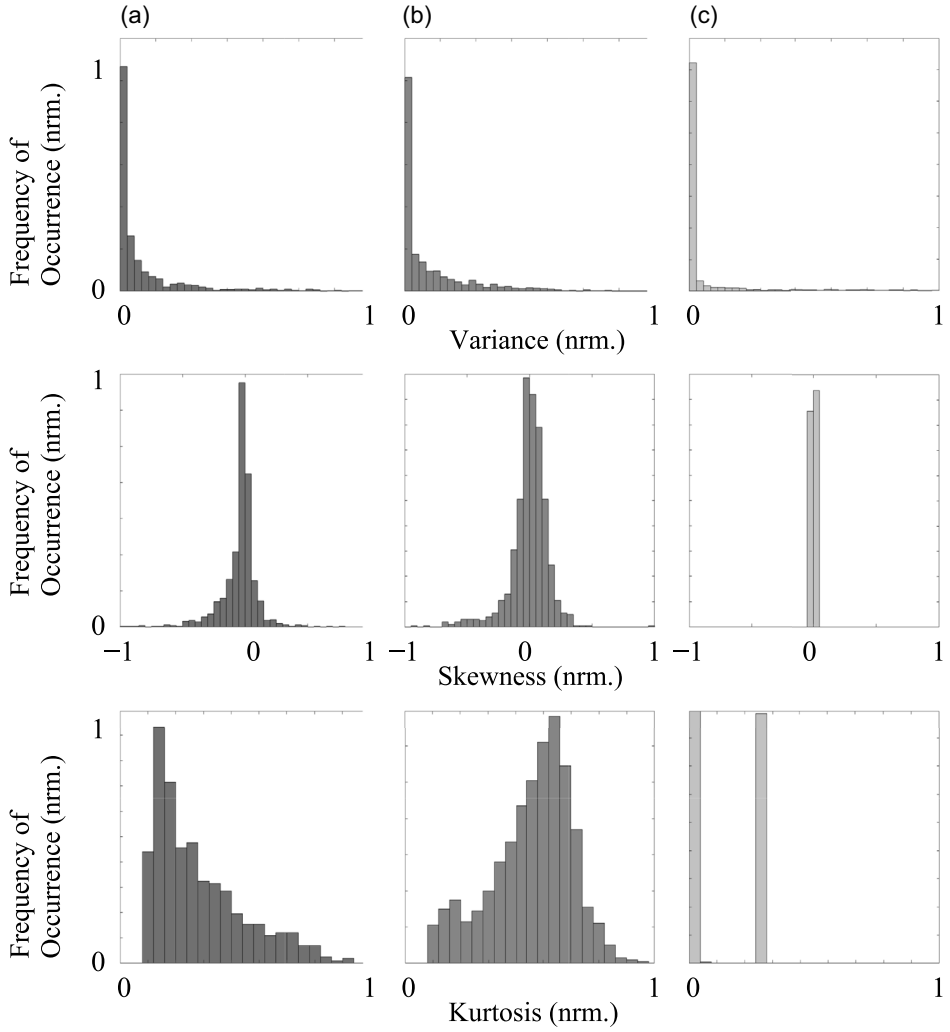

Figure 7. Standardized moments of the amplitudes in the measured and generated impulse responses. Histograms of 1000 samples taken from each dataset. Histograms show the distributions of the moments from column a) cVAE-GAN generated IRs, column b) measured IRs, and column c) IRs obtained from the disk model.

had a training accuracy of ±3.8° and a test accuracy of ±5.3°. The receiver operating characteristic (ROC) of the leaf-species classifier had an area under the curve (AUC) (averaged over all four leaf species) of 0.98.

After the cVAE-GAN system was trained on the measured dataset, impulse responses were generated to be compared to the measured data qualitatively. In these comparisons, substantial similarity was observed between measured and generated impulse responses (**Figure 6**): The general shape and pattern of the impulse responses were observed to be similar. At first it appeared that there was more noise present in the generated IRs compared to the measured IRs, but after taking the standard deviation (σ^2 of the first 50 samples of each IR in the measured dataset (where the meaningful signal is less likely to be present), and taking the σ^2 of the first 50 samples of generated IRs conditioned by the same labels as the measured data, there was found to be less noise in the generated IRs compared to the measured (the mean σ^2 for measured data was 0.00152, while for generated it was 0.00051, the higher level in the measured data being due to more outliers).

The first quantitative comparison was performed by computing the standardized moments of the waveform amplitude values. All three standardized moments (variance, skewness, and kurtosis) of the impulse responses were more similar between the cVAE-GAN model and measured data, compared to the previous state of the art disk model and real data (**Figure 7**). Taking the distributions of these moments, the KL divergence was computed. The KL divergence between the variances of the measured and the cVAE-generated IRs was 0.10, while the KL divergence between the measured IRs and those obtained from the disk model was 1.91, i.e., 19 times larger than for the cVAE-generated IRs. For the distribution of skewness, the KL divergence for measured vs. cVAE-generated was 0.18, while for the comparison with the disk-model IRs it was 0.64, i.e., 3.6 times larger than for the cVAE-generated IRs. Finally, the Kl divergence of the distributions of the kurtosis was 0.77 for the cVAE-generated IRs, compared to 3.94 for the disk model, i.e., the divergence was 5 times larger for the disk model.

Next, a total of 10 000 IRs were generated with the cVAE-GAN for a variety of azimuth angles ranging from 0° to 90°. The total signal energy (represented by a sum of squares of the signal amplitudes) of the generated IRs varied with azimuth angle in a similar way to what was seen in the measured IRs. There was a peak at 0°, a 'hump' between 0°and 45°, and a flatter energy level between 45° and 90° (**Figure 8**).

Two regressors were then trained to predict azimuth angles from the IRs, one was trained on cVAE-generated IRs and the other trained on the measured IR dataset. The regressor trained on the cVAE-generated IRs then predicted the azimuth angle of both cVAE-generated and measured IRs. The regressions showed similarly low errors for both test cases. The mean test error for the cVAE-generated IRs was 11.7°, while the test error

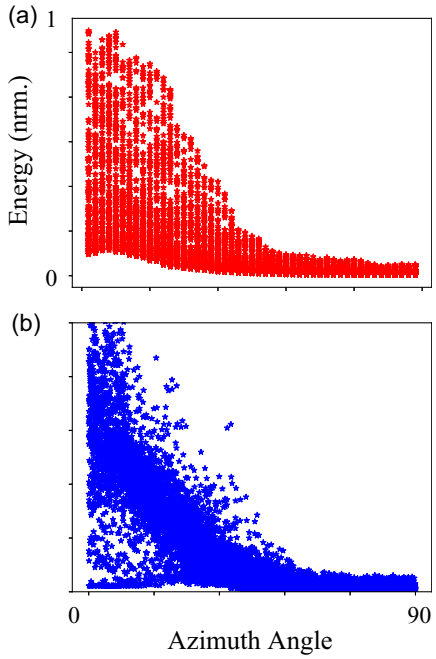

Figure 8. Signal energy (estimated as the sum of the squared signal amplitude) in the IRs as a function of the azimuth angle under which the leaf is being ensonified. a) Measured IRs and b) cVAE-GAN-generated IRs.

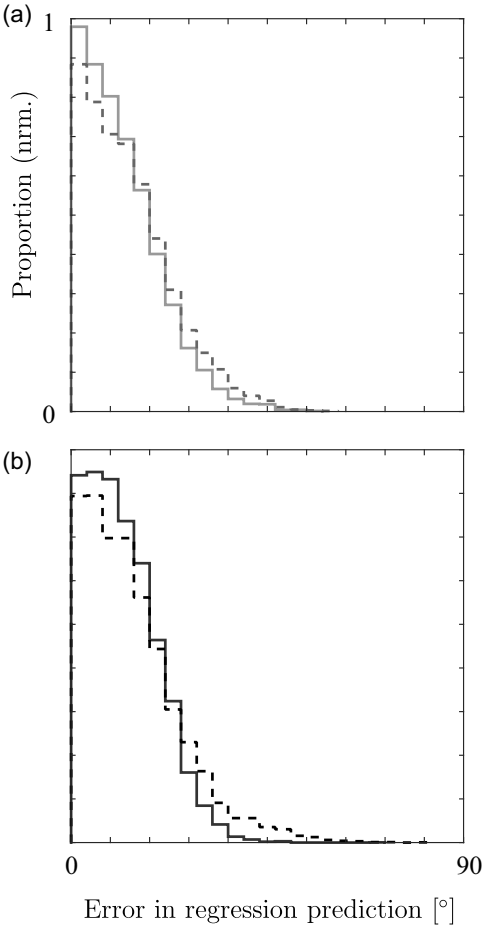

Figure 9. Distribution of regression errors for the estimation of azimuth angle. a) Errors of a regressor trained on cVAE-GAN-generated data and evaluated on cVAE-GAN-generated data (solid line) and measured data (dashed line). b) Errors of a regressor trained on real data and evaluated on real data (solid line) and fake data (dashed line).

for real IRs was 13.2°. The regressor trained on the real IRs which then predicted the azimuth angle of both generated and real IRs similarly showed low errors and similar amounts of error for both test cases (**Figure 9**). The mean test error for the generated IRs was 14°, while the test error for real IRs was 11.9°.

4. Conclusion

The results of the current study demonstrate that a VAE-GAN-based generative method is capable of producing synthetic impulse responses that mimic those of leaves: For a human observer, it was very difficult to decide whether any of the signals in this study was measured or generated. There were no obvious distinguishing features in the duration, magnitude, and waveform shape of the signals. The presence of noise in VAE generated images is a known perennial problem with many VAE-based methods,[22] but by comparing the noise levels at the start of the signals we have shown that this is not an issue for our model.

The finding that classification by leaf species and regression by azimuth angle were both readily possible (Figure 5) demonstrates that the measured impulse responses contained this kind of information. A good foliage model should hence also reproduce the influence of these variable on the acoustic characteristics of the leaves. The results from the current study show that a cVAE-GAN generator is capable of achieving this. The result that a regressor trained on the measured data and tested on the cVAE-GAN-generated data, and vice versa (Figure 9), produced low regression error provides solid evidence that our generation method is incorporating this conditional information into the generated IRs. The same argument can be made for the similarity of the energy levels as a function of viewing azimuth angle (Figure 8) as energy was not explicitly given to the generator, but had to be learned implicitly.

The first three standardized moments of the signal amplitude distributions for signals produced by our method were much closer to the respective parameters of the measured data than was the case for the previous state of the art (Figure 7). This can be taken as quantitative evidence that our method generates more realistic impulse responses than what had been achieved previously. The moments were not provided as training information explicitly, but had to be learned indirectly through the training process. The similarity that was achieved demonstrates that this learning has taken place. However, since the distribution of the kurtosis for the cVAE-GAN-generated data was less close to the value obtained for the measured IRs than was the case for the second and third moments (although still more similar than the previous state of the art), there is clearly room for improvement, and future methods or modifications to our method may be able to generate IRs that are even more similar to measured IRs in this and perhaps other respects.

Our generative leaf model, being based on brief impulse responses, is computationally efficient enough to recreate large-scale acoustic environments of trees. From the qualitative inspection and the moment analysis conducted here, it also appears that the echoes would be similar enough to real foliages in their statistical properties and how vary as a function of parameters such as leaf species and viewing angle to make experimenting with such a virtual environment worthwhile. Like the prior state of the art, our model follows the Born approximation (ignores multiple bounces between leaves) and ignores shadowing of deep leaves from shallow leaves. While these effects certainly exist, and we believe them to be of minimal importance, future work may take these into account when using deep learning based generated impulse responses to implement full foliage environment simulations.

These environments would be useful for modelling bat biosonar navigation strategies and aid in engineering biomimetic sonar devices to add new sensing modalities for autonomous drones in dense foliage settings. Hopefully this work will inspire generative machine learning approaches to the simulation of other complex signals in domains such as underwater sensing, radar, and medical ultrasound.

Any application of deep generative modelling methods cannot be an exhaustive search of all methods and all possible settings of hyper-parameters, thus it is likely that superior modelling using a tweaked version of this method or an entirely different architecture may be have results superior to ours, especially given our failure to perfectly capture the distribution of kurtosis values of real IRs. A principled approach to find such a superior method is an interesting open question in machine learning.

Since our model does not correspond to any particular realisation of foliage geometry, a comparison with numerical methods based on individual leaf echoes, such as finite-difference time-domain (FDTD) or finite element methods (FEM), is not feasible. We have used real leaf echoes for comparison to our method, which is the ultimate measure of success.

Acknowledgements

Thank you to the NSF CDS&E (Award ID 1762577) and NAVSEA/NEEC (grant no. N001742210007) for funding this work. Also thank you to Lucas Mun and Nathan Cox for assisting in setting up the data acquisition system.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available in [Leaf impulse responses] at [https://doi.org/10.6084/m9.figshare.24437920.v1], reference number [24437920].

Keywords

bioacoustics, deep learning, variational autoencoders

Received: October 26, 2023
Revised: January 2, 2024
Published online:

[1] A. S. Aguiar, F. N. dos Santos, J. B. Cunha, H. Sobreira, A. J. Sousa, *Robotics* **2020**, *9*, 4.

[2] L. Wallace, A. Lucieer, C. Watson, D. Turner, *Remote Sens.* **2012**, *4*, 1519.

[3] C. Wang, J. Wang, Y. Shen, X. Zhang, *IEEE Trans. Veh. Technol.* **2019**, *68*, 2124.
[4] G. Neuweiler, *Trends Ecol. Evol.* **1989**, *4*, 160.
[5] J. A. Thomas, C. F. Moss, M. Vater, *Echolocation in Bats and Dolphins*, University of Chicago Press, Chicago, IL **2004**.
[6] R. Müller, R. Kuc, *J. Acoust. Soc. Am.* **2000**, *108*, 836.
[7] W. S. Burdic, *J. Acoust. Soc. Am.* **1991**, *89*, 3020.
[8] B. D. Todd, R. Müller, *Bioinspiration Biomimetics* **2017**, *13*, 016014.
[9] P. Zhu, J. Isaacs, B. Fu, S. Ferrari, in *2017 IEEE 56th Annual Conf. Decision and Control (CDC)*, IEEE, Piscataway, NJ **2017**, pp. 2724–2731.
[10] R. R. Wildeboer, F. Sammali, R. J. G. van Sloun, Y. Huang, P. Chen, M. Bruce, C. Rabotti, S. Shulepov, G. Salomon, B. C. Schoot, H. Wijkstra, M. Mischi, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2020**, *67*, 1497.
[11] P. McKerrow, N. Harper, *IEEE Sens. J.* **2001**, *1*, 245.
[12] Y. Yovel, P. Stilz, M. O. Franz, A. Boonman, H.-U. Schnitzler, *PLoS Comput. Biol.* **2009**, *5*, 1000429.
[13] C. Ming, A. K. Gupta, R. Lu, H. Zhu, R. Müller, *PLoS One* **2017**, *12*, 0182824.
[14] U. Taskin, N. Ozmen, H. Gemmeke, K. W. Van Dongen, *Arch. Acoust.* **2018**, *43*, 425.
[15] P. Belanger, P. Cawley, F. Simonetti, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2010**, *57*, 1405.
[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, in *Advances in Neural Information Processing Systems* (Eds: Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, K. Q. Weinberger), Curran Associates, Inc., Red Hook, NY **2014**, p. 27.
[17] D. P. Kingma, M. Welling (Preprint), arXiv:1312.6114, v11, submitted: Dec. **2013** .
[18] D. Saxena, J. Cao, *ACM Comp. Surv.,* **2021**, *54*, 63.
[19] H. Alqahtani, M. Kavakli-Thorne, G. Kumar, *Arch. Comput. Methods Eng.* **2021**, *28*, 525.
[20] S. Dieleman, A. van den Oord, K. Simonyan, in *Advances in Neural Information Processing Systems* (Eds: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett), Curran Associates, Inc., Red Hook, NY **2018**, p. 31.
[21] C. Donahue, J. McAuley, M. Puckette (Preprint), arXiv:1802.04208, v3, submitted: Feb. **2018**.
[22] R. Wei, C. Garcia, A. El-Sayed, V. Peterson, A. Mahmood, *IEEE Access* **2020**, *8*, 153651.
[23] A. Singh, T. Ogunfunmi, *Entropy* **2022**, *24*, 55.
[24] M. F. Mathieu, J. J. Zhao, A. Ramesh, P. Sprechmann, Y. LeCun, in *Advances in Neural Information Processing Systems* (Eds: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett), Curran Associates, Inc., Red Hook, NY **2016**, p. 29.
[25] Z. Zheng, L. Sun, in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA **2019**.
[26] J. Bao, D. Chen, F. Wen, H. Li, G. Hua, in *2017 IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, Italy **2017**.
[27] A. Asperti, M. Trentin, *IEEE Access* **2020**, *8*, 199440.
[28] D. P. Kingma, J. Ba (Preprint), arXiv:1412.6980, v9, submitted: Dec. **2014**.