A Variational Auto-Encoder Model for Underwater Acoustic Channels

Li Wei and Zhaohui Wang
Michigan Technological University
Houghton, Michigan, USA
liwei@mtu.edu,zhaohuiw@mtu.edu

Abstract

An underwater acoustic (UWA) channel model with high validity and re-usability is widely demanded. In this paper, we propose a variational auto-encoder (VAE)-based deep generative model which learns an abstract representation of the UWA channel impulse responses (CIRs) and can generate CIR samples with similar features. A customized training process is proposed to avoid the model collapse and being trapped in a gradient pit. The proposed deep generative model is validated using field experimental data sets.

CCS Concepts: • **General and reference → General conference proceedings**.

Keywords: Generative model, variational autoencoder, underwater acoustic communication

ACM Reference Format:
Li Wei and Zhaohui Wang. 2021. A Variational Auto-Encoder Model for Underwater Acoustic Channels. In *The 15th International Conference on Underwater Networks Systems (WUWNet'21), November 22–24, 2021, Shenzhen, Guangdong, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3491315.3491330

1 Introduction

There are an increasing number of underwater systems that have recently emerged for underwater exploration, offshore resource extraction, aquaculture, environmental monitoring, oceanographic research, national defense missions, etc. As various mobile systems have been introduced to these underwater applications, the demands for underwater acoustic (UWA) communications and networking are growing rapidly. To deal with the challenging UWA channel conditions, a rapid growth of novel UWA communication schemes have drawn much attention in recent years. The study of UWA channel characteristics plays an important role through the design, analysis, and performance evaluation processes of these novel UWA communication schemes. Therefore, a UWA channel model with high validity and re-usability is widely demanded. Such an effective model is expected to have several layers of complexity that can reflect the deterministic characteristics of acoustic propagation in a certain geographical environment, as well as the stochastic characteristics caused by large-scale and small-scale uncertainties [10].

The UWA channel characteristics have been studied with both modeling and experimentation methodologies for decades [1]. UWA channels have been represented as mathematical models or simulation models. The mathematical models describe the variations of an acoustic waveform when propagating through a UWA channel, which play an important role in the design and analysis of UWA communication systems. However, tractable mathematical descriptions of an UWA communication channel are elusive due to the complexity that how environmental parameters affect the sound propagation, reflection, refraction, scattering and reverberation. The simulation models can generate UWA channel realizations based on mathematical models describing the variation of acoustic waveforms [1], stochastic models describing the distributions of UWA channel randomness [8], and/or replays of the measured channel condition in experiments [7]. Most existing simulation systems can simulate certain aspects of the UWA channel, but few systems have demonstrated the capability of simulating the UWA channel which can match the data over a time scale that is appropriate for UWA communications [14].

On the other hand, performance evaluation of practical UWA communication systems still heavily relies on extensive field experiments due the lack of well-categorized channel types and corresponding stochastic models. Although several field experiment test-beds have been developed in past years, the field experiments are still costly and have limited opportunity for repeating tests [12]. The re-usability of field experiment data is limited since field experiments are usually tailored to a particular communication scheme. Even for the same communication scheme, data obtained at different experiment sites could be distinctive due to the geographic and hydro-graphic differences of UWA channels. Thus, a UWA channel model with high validity and re-usability is widely demanded due to the limitations of existing mathematical models, simulation models and field experiments.

WUWNet '21, November 22-24, 2021, Shenzhen, Guangdong, China

ACM ISBN 978-1-4503-9562-5/21/11…$15.00
https://doi.org/10.1145/3491315.3491330

Deep learning has been found a wide range of applications in wireless communication systems in recent years. The Deep generative models [2] can learn from a target data distribution and generates new samples following a similar distribution, which function in a similar way as that of the UWA simulation channel models and play-back models, but no prior knowledge of stochastic characteristics or mathematical models is needed. The variational auto-encoders (VAE) [5] can be promising for modeling UWA channels since both the deterministic and stochastic characteristics of UWA channels could be distinctive for different scenarios even for different cases or at different time in the same water body. For example, a water body could be covered with ice in the winter, and the surface reflection of acoustic waveform with different types of ice cover conditions will significantly affect the UWA channel characteristics. Depending on the speed of temporal variations, these affects on the UWA channel can be considered as changes of the deterministic characteristics when the ice cover condition varies slowly, or changes of the stochastic characteristics for fast ice cover condition variations. Since the VAE will learn from the observations of the UWA channel, no prior knowledge about the stochastic characteristics is needed for modeling the UWA channel, and the UWA channel deterministic characteristics can also be modeled without an explicit mathematical description of the ice cover condition. New UWA channel data generated from this model can be used to evaluate the performance of a certain UWA communication scheme, as well as to train deep learning models for other applications which require a large training data set.

Several researchers have explored utilizing auto-encoder based models for solving various wireless communication problems. The encoder and decoder perform signal processing at the transmitter and the receiver, respectively. Traditional digital signal processing modules, such as error correction coding, components of the modulation and demodulation, and detection, are implemented as the encoder at the transmitter and as the decoder at the receiver. The auto-encoder is trained as parts of a communication system including a transmitter, a receiver, and a channel model in between. Other signal processing modules at the transmitter and receiver side are also included in the overall input-output of the auto-encoder, as well as the distortions and noises added to the signal by the channel. Existing auto-encoder based communication system design work has been reviewed in [9] and [15]. However, most existing works suffer from the curse of dimensionality and can only be evaluated with a simple additive white Gaussian noise (AWGN) channel model. Another type of the auto-encoder application for wireless communication is that using an auto-encoder to compress the downlink channel state information (CSI) of multiple-input multiple-output (MIMO) wireless communication systems to reduce the CSI feedback overhead. An optional channel model can also be involved in the training process of the auto-encoder to enhance the overall robustness of the compressed MIMO CSI feedback. Related works for this application have been summarized in [3]. An auto-encoder consisting of residual network blocks was employed in [13] to reduce the dimensionality of CSI representations. Furthermore, an extended neural network structure with learning rate scheduling scheme was proposed in [6] to enhanced the MIMO CSI compression performances. Both works demonstrated practical scales and architectures of deep neural networks for representation learning of CSI data, which have significant referential value for VAE design of CSI generative models.

In this paper, we propose a deep generative model for the UWA CIR based on VAE. Field experimental data sets are employed to train the proposed generative model. Existing auto-encoder models for the CIR, namely CsiNet [13] and CRNet [6], are evaluated and modified to VAE. The training process of such models are customized to prevent the VAE model collapses and ensure that a practical generative process could be performed.

2 A generative model for UWA channels

The UWA channel can be modeled as a channel impulse response (CIR) $\boldsymbol{h}$. Given the transmitted signal $\boldsymbol{x}$, the channel output can be formulated as a convolution between the CIR and the input,

$$\boldsymbol{y} = \boldsymbol{h} * \boldsymbol{x} + \boldsymbol{n} \tag{1}$$

with $\boldsymbol{n}$ being the channel noise. The CIR $\boldsymbol{h}$ can be considered as following a stochastic distribution conditioning on the environmental parameters of the water body and channel dynamics. The proposed generative model is to learn from a set of the $\boldsymbol{h}$ observations, then generate new CIR samples following a similar distribution.

2.1 VAE as generative models

An auto-encoder model consists of an encoder and an decoder. The encoder converts a data sample to an abstract representation in a latent hyperspace, while the decoder can convert a sample in the latent space back to a data sample in the original form. With constraints on the latent space, the abstract representations of the original data sample can have different number of dimensions and disentangled correlations among latent dimensions. Instead of converting the original data into a point in the latent hyperspace, the encoder of a VAE converts a original data sample into a set of parameters describing a stochastic distribution in the latent space, and a reparameterization module is introduced between the encoder and the decoder which draws a sample from the distribution determined by the encoded set of parameters each time when decoding.

The loss function of a VAE consists of the reconstruction error and several regularization terms. The reconstruction error is introduced to minimize the difference between the

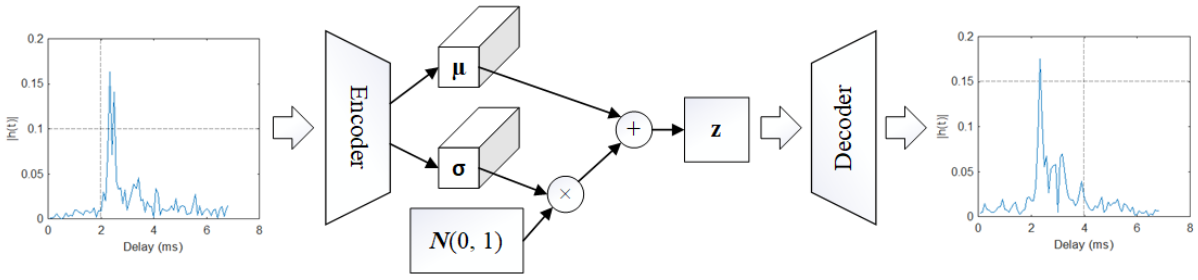

Figure 1. Overview of the deep generative model for underwater acoustic channel impulse responses (CIRs)

decoded output and the the input. The regularization terms reduce the correlation of latent dimensions and also constrain the distributions of encoded representations to be similar to given prior distributions. By sampling from the prior distributions in the latent space, new data samples sharing similar deterministic and stochastic characteristics of the original data can be generated with the trained decoder of the VAE.

2.2 The VAE model for UWA CIRs

An overview of the deep generative model is shown in Fig. 1. The input $\boldsymbol{h}$, taking from a field experiment, is originally a vector consisting of L complex values, and its real and imaginary parts are first reshaped to a $(2, \sqrt{L}, \sqrt{L})$ real value tensor. The encoder converts $\boldsymbol{h}$ to two parameter vectors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. The latent representation $\boldsymbol{z}$ is obtained from the reparameterization process based on the parameter vectors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. The decoder takes samples of the latent representation $\boldsymbol{z}$ as input and finally outputs the CIR reconstruction $\boldsymbol{h'}$.

The loss function of the VAE consists of the reconstruction error and a standard VAE regularization term. The mean squared error (MSE) of the reconstruction and the the original input is employed as the reconstruction error.

$$\min \mathcal{L} = \mathrm{MSE}(\boldsymbol{h}, \boldsymbol{h'}) + \frac{1}{2}\sum_{j=1}^{M}\left(1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2\right) \quad (2)$$

where M is the dimensionality of latent representation $\boldsymbol{z}$ in latent hyperspace, and $\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \cdot \boldsymbol{\epsilon}$, where vector $\boldsymbol{\epsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I})$ is drawn from a multivariate normal distribution of dimension M with an identity covariance matrix, and $\cdot$ denotes element-wise product.

Multiple network architectures are available to construct the encoder and the decoder. The CsiNet[13] and CRNet[6] were both validated workable auto-encoder architectures but without the reparameterization module. Other typical network architectures such as DenseNet[4] or classic simple CNNs can also be employed to construct the deep neural network. The neural network with insufficient depth and width will result in inaccurate similar reconstruction which is more similar to an average of all data samples rather than distinguishing reconstructions of each different data samples. However, neural networks with too many parameters suffer from the curse of dimensionality. Here as an demonstration, we employ the neural network architecture of CsiNet[13] and introduce the reparameterization module of the VAE. The encoder consists of 2 CNN layers and a linear layer. For a CIR of length $L = 256$, this leads to a total dimensionality of 16 in the latent space. The decoder consists of a linear layer that increases the dimensionality and then 2 RefineNet blocks proposed in [13]. There are 28,472 parameters in total to be trained.

2.3 Training process of the VAE for UWA CIRs

Considering the very large dimensionality of $\boldsymbol{h}$, if the proposed model is trained following the classic VAE training process, the model will easily collapse and its parameters will be trapped in a gradient pit. A typical phenomenon is that no matter whatever inputs to the encoder, its output $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the same and very close to zero vectors. Another typical phenomenon is that all the reconstructions output from the decoder are the same no matter whatever the value of $\boldsymbol{z}$ is.

To obtain a practical VAE for generating CIR samples, the training process can be divided into 3 stages. First, train the parameters as a basic auto-encoder: remove the regularization terms from the loss function and use only the MSE of reconstruction value for gradient back propagation. The reparameterization process should also be disabled and let $\boldsymbol{z} = \boldsymbol{\mu}$ during this stage. Reasonable reconstruction that is similar to the input can be obtained during this stage. Second, when the reconstruction error is small enough, introduce the regularization terms into the loss function but still disable the reparameterization process. A threshold can be set up, the training can be switched between Stage 1 and 2 based on whether the reconstruction error is below the threshold. During this stage, the $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ will gradually move towards a zero vector $\boldsymbol{0}$ and $\boldsymbol{I}$, respectively. Finally, when $\boldsymbol{\sigma}$ is stabilized and close to vectors with very small amplitudes and the reconstruction error is bouncing around the threshold, the reparameterization process can be enabled, but the regularization terms in the loss function can be disabled. The

second and the third stage can only take a few epochs as long as the $\boldsymbol{\mu}$'s of most training data set distribute closely to a multivariate standard normal distribution, while the reconstruction $\boldsymbol{h'}$ is similar to the input and distinguishable from each other.

The model collapse could also happen with such training configuration. Thus, saving checkpoints of the model parameters with the current best loss value and resuming the model parameters to a checkpoint are both necessary.

3 Experiments and Result Analysis

The proposed model are evaluated with field experiment data sets KWAUG14[11], which was obtained in an experiment conducted in the Keweenaw Waterway. The water depth was around 3-5m. The acoustic modems were at 1.5m depth and the transmission distance was 313m. The CIR length is $L = 256$. The batch size is set to 50, and there are 774 batches of training data, 86 batches of validation data. The histogram of CIR amplitudes in KWAUG14 is showing in Figure 2.

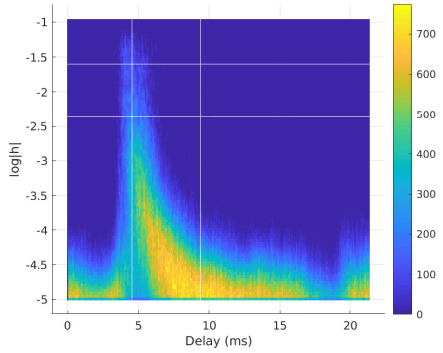

Figure 2. Histograms of the KWAUG14 CIR amplitude

The CsiNet's architecture is employed as the encoder and decoder. The learning rate is set to 0.002 and Adam optimizer is employed for training. Figure 3 shows the loss value at each epoch for the training process. The Stage 2 begins around Epoch 1600, and the Stage 3 begins around Epoch 1800.

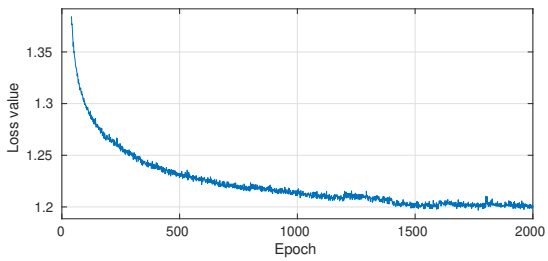

Figure 3. Loss value during training

3.1 Reconstruction of CIR samples

Figure 4 shows 3 samples of the KWAUG14 CIR amplitude and their corresponding reconstructions with the trained VAE. The reconstruction is obtained with the reparameterization process disabled. The results shows that significant values of CIR are successfully captured in the reconstruction.

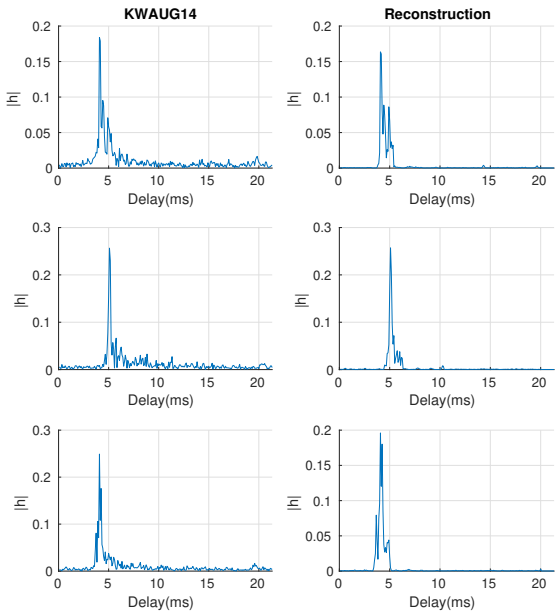

Figure 4. Samples of the KWAUG14 CIR and corresponding reconstructions

The histogram of all reconstructions is shown in Figure 5 (right). The significant values of CIRs locate in the similar area as in Figure 2.

3.2 Distribution of the latent representations

The histogram of the encoded output $\boldsymbol{\mu}$ is shown in Figure 5 (left). Although the latent representations do not strictly follow the normal distribution in all latent dimensions, most latent dimensions are centered around 0 and have similar variances.

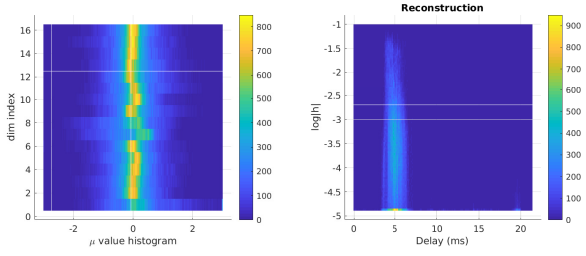

Figure 5. Histograms of latent representations and reconstructions

3.3 Generated CIR samples and distribution

The mean and variance of the latent representation in each dimension in Figure 5 (left) are calculated. To generate new CIR samples, we first draw in each dimension 10,000 random numbers following a normal distribution that has the same mean and variance as shown in Figure 5 (left). The distribution of (10000, 16) random numbers is shown in Figure 6 (left). Then, these random numbers are processed by the trained decoder and new CIRs can be obtained. The histogram of generated CIRs are showing in Figure 6 (right). The histogram is similar to that of the reconstruction histogram in Figure 5 (right).

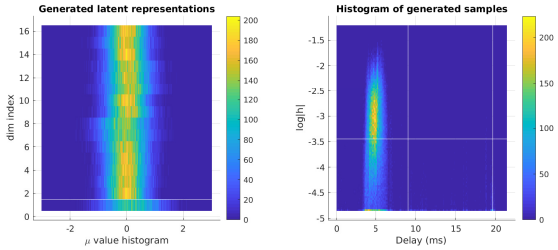

Figure 6. Histograms of generated latent representations and corresponding CIRs

Figure 7 shows 4 generated CIR samples. They share the similar features as those shown in Figure 4.

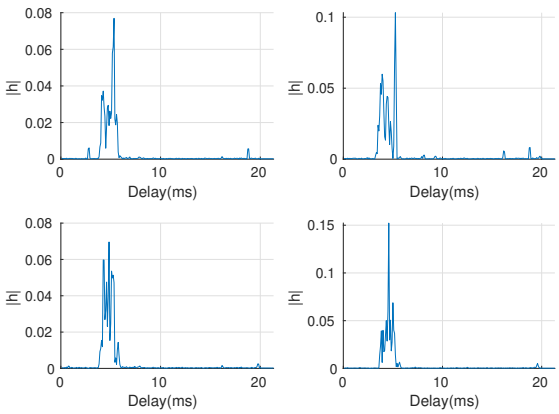

Figure 7. Samples of generated CIRs

4 Conclusions

In this paper, we explored utilizing VAE to learn a generative model for the UWA CIR. Due to the dimensionality of the CIR data set, the classic training process of VAE is not practical. A modified training process is proposed. Results show that the reconstructions and generated data have satisfying performance.

Acknowledgments

This work was supported by the NSF grant ECCS-1651135 (CAREER).

References

[1] Paul C Etter. 2018. *Underwater acoustic modeling and simulation.* CRC press.

[2] David Foster. 2019. *Generative deep learning: Teaching machines to paint, write, compose, and play.* O'Reilly Media.

[3] Jiajia Guo, Chao-Kai Wen, Shi Jin, and Geoffrey Ye Li. 2020. Convolutional neural network-based multiple-rate compressive sensing for massive MIMO CSI feedback: Design, simulation, and analysis. *IEEE Transactions on Wireless Communications* 19, 4 (2020), 2827–2840.

[4] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 4700–4708.

[5] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

[6] Zhilin Lu, Jintao Wang, and Jian Song. 2020. Multi-resolution CSI feedback with deep learning in massive MIMO system. In *ICC 2020-2020 IEEE International Conference on Communications (ICC).* IEEE, 1–6.

[7] Roald Otnes, Paul A van Walree, and Trond Jenserud. 2013. Validation of replay-based underwater acoustic communication channel simulation. *IEEE Journal of Oceanic Engineering* 38, 4 (2013), 689–700.

[8] Parastoo Qarabaqi and Milica Stojanovic. 2013. Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels. *IEEE Journal of Oceanic Engineering* 38, 4 (2013), 701–717.

[9] Zhijin Qin, Hao Ye, Geoffrey Ye Li, and Biing-Hwang Fred Juang. 2019. Deep learning in physical layer communications. *IEEE Wireless Communications* 26, 2 (2019), 93–99.

[10] Aijun Song, Milica Stojanovic, and Mandar Chitre. 2019. Editorial Underwater Acoustic Communications: Where We Stand and What Is Next? *IEEE Journal of Oceanic Engineering* 44, 1 (2019), 1–6.

[11] Wensheng Sun and Zhaohui Wang. 2018. Online modeling and prediction of the large-scale temporal variation in underwater acoustic communication channels. *IEEE Access* 6 (2018), 73984–74002.

[12] Li Wei, Zheng Peng, Hao Zhou, Jun-Hong Cui, Shengli Zhou, Zhijie Shi, and James O'Donnell. 2013. Long island sound testbed and experiments. In *2013 OCEANS-San Diego.* IEEE, 1–6.

[13] Chao-Kai Wen, Wan-Ting Shih, and Shi Jin. 2018. Deep learning for massive MIMO CSI feedback. *IEEE Wireless Communications Letters* 7, 5 (2018), 748–751.

[14] T. C. Yang and S. H. Huang. 2016. Building a database of ocean channel impulse responses for underwater acoustic communication performance evaluation: issues, requirements, methods and results. In *Proceedings of the 11th ACM International Conference on Underwater Networks & Systems.* 1–8.

[15] Hao Ye, Le Liang, and Geoffrey Ye Li. 2019. Circular convolutional auto-encoder for channel coding. In *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC).* IEEE, 1–5.