On the use of Generative adversarial neural networks for computing photonic crystal fiber optical properties

Zelaci Aimen, Ahmet Yaşlı, Cem Kalyoncu, and Hüseyin Ademgil

Abstract

Photonic crystal fibers (PCF) for specific applications are designed and optimized by both industry experts and researches. However, the potential number of combinations possible for a single application is huge. This issue combined by the speed of PCF numerical simulation techniques causes the task to take significant amount of time. As stated in the previous works, artificial neural networks (ANN) can predict the result of numerical simulations much faster. However, there are two issues with the methods proposed previously. Namely, the required number of samples for training and generality of the designed network. In this paper, we have proposed the use of generative adversarial networks (GAN) to fabricate additional data to be used in training and a unique ANN design that can work with wider range of configurations.

I. INTRODUCTION

Importance of PCF and SPR, written by AY or HA

Importance of ANN, deep neural networks, GAN

Literature survey

Short explanation of GAN/ANN

This paper is organized as follows. Section II details the use of GAN to generate additional training samples for ANN as well as the proposed neural network architecture. Photonic crystal fiber design that is used for testing is described in details in section III. Detailed analysis of the experimental results are discussed in section IV. Finally, concluding remarks are made in section V.

II. PROPOSED METHOD

Details of the overall architecture

A. Artificial neural network design

The ANN model is a fully-connected feed-forward MLP (Multi Layer Perceptron), consisting of an input layer, an output layer, and 5 hidden layers, 50 neurons for each hidden layer. Taking ReLU (Rectified Linear Unit) as the activation function, Adam (Diederik P. Kingma, Jimmy Ba, 22 Dec 2014) as the optimizer and the mean suqared error as the loss/cost function. To reduce overfitting we use the Early stopping method [reference], that is by saving

checkpoints where the best validation mean squared error occurred as we iterate. To accelerate training and mitigate the problem of internal covariate shift Batch Normalization algorithm was used [reference].

B. Generative Adversarial networks

GANs consist of two neural networks, as shown in figure (3). The first of which called a discriminator, that gives an estimate of the probability that a given input is real or generated (fake). Whereas the second network is referred to as the generator, which outputs data samples from a random noise vector called a latent variable usually given the symbol z supplied at its input. The error between the discriminator's output and the actual labels (The real data samples vectors all labeled as 1: the probability of being real) would then be measured by the means of a chosen metric. Introduced by Arjovsky et al [reference] the Wasserstein distance metric (or the earth mover distance) proved to be very effective, instead of discriminating whether an input is real or generated, the discriminator provides a criticism of how far the generated data from the real data is, hence the discriminator network is referred to as the critic in the WGANs. Throughout this paper we will use yet the improved WGAN, the WGAN with Gradient penalty [reference]. The improved WGAN converges in a stable manner with least efforts made to tune the hyperparameters of both networks constituting the GAN, which can be very daunting. GANs power emerges from the fact that discover insights or structures embedded within the data, which enables it generate it samples within an acceptable error. Therefore, in this work we will try to supplement our original training with the generated samples to improve the ANN model accuracy on predicting the confinement loss of the PCF.

III. PHOTONIC CRYSTAL FIBER DESIGN

PCF design details, written by AY

IV. EXPERIMENTS

A. Experimental setup

A labeled data-set of only 432 samples was collected in this work, through simulations using FV-FEM. The length of the data-set made the task of building ANN model look quiet impossible, however the results were satisfying to some extent. The data-set consists of the wavelength λ, index of refraction $n_{analyte}$, air-hole to air-hole distance Λ, and the air holes radii per ring $d1$, $d2$ and $d3$, taken as our independent variables. The labels are the confinement loss of the PCF. The set consists of nine different configurations of the geometric properties (Λ, d1, d2, d3), for each configuration the confinement loss was calculated for 3 different analytes (Water (n=1.33), Ethanol (n=1.35) and several commercial hydrocarbon mixtures (n=1.36) [35]). Seven configurations were randomly selected for training both the WGAN-GP and the ANN, one configuration was held out for validation, and the last configuration for testing. This data was preprocessed before fed in the networks. The indices of refraction are very close, which made it quiet difficult for the neural networks to differentiate between them, after many trials, the best choice was to take only the tens digit of $134, 135, 136$, giving $4, 5, 6$. Finally, we transform the confinement loss to the log scale.

Metrics used in the comparisons

B. Performance of ANN

We train the ANN model for more than 2000 epochs, starting from the original data-set. Then we augment the data by 1000 generated samples from our WGAN-GP, which will be demonstrated in the next section. The following table summarizes the hyper-parameters chosen for the ANN model:

Length of the training data-set	Learning rate	Batch size	β_1	β_2
336	1×10^{-04}	12	0.9	0.999
1000+336	1×10^{-04}	12	0.9	0.999
2000 + 336	2×10^{-04}	16	0.9	0.999
3000 + 336	2.5×10^{-04}	16	0.9	0.999

The MSE on the training sets ranged from 0.0030 to 0.0050. For all data sets the MSE on the training sets decreased in an acceptable manner, as shown in the following figure.
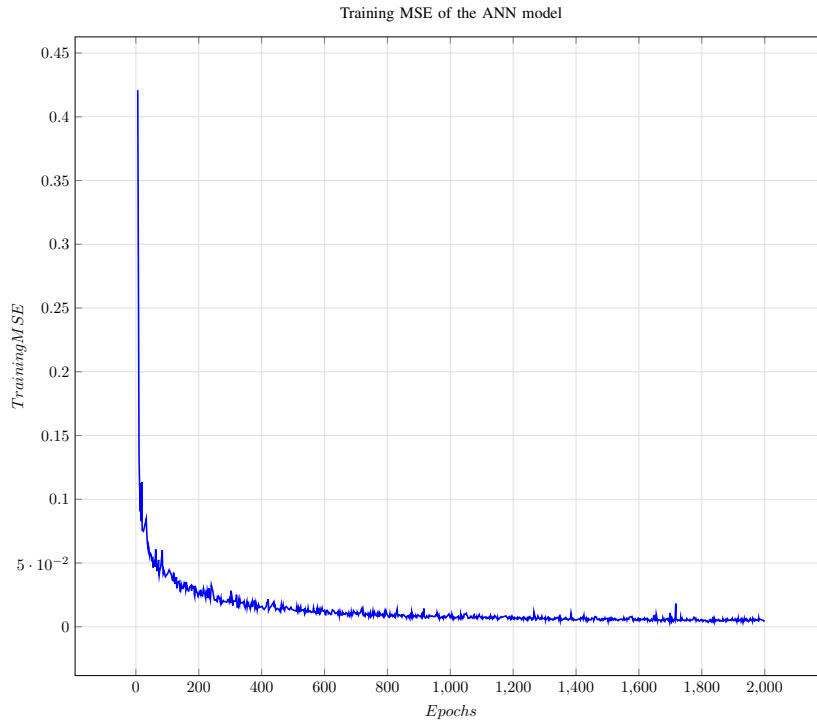

Figure 1.

Next, we plot the predictions the ANN model made after training using only the original data-set.
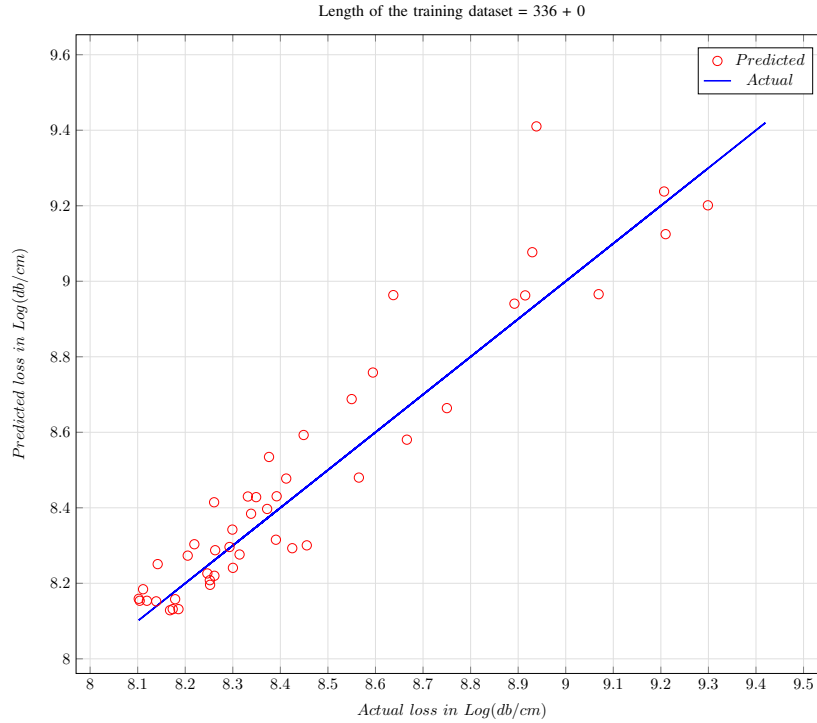
Length of the training dataset = 336 + 0

Figure 2.

As you can see, the ANN made somewhat good predictions of the low values of the confinement loss. A close inspection into the data samples collected shows that points are concentrated in the low values interval, i.e the ANN picked the pattern quiet easily. However in the higher values interval, the ANN could not see a pattern because points are distant from each other to some extent, even though we transformed the confinement loss to the log scale to actually minimize these distances.

C. Performance boost of GAN

Experiments regarding to the GAN

D. Computational performance

Training and execution time of ANN versus simulation method

V. CONCLUSION

about the improved speed

reduced amount of training samples

increased generality of the system

Machine learning approaches has an inherit strength on top of classical simulation methods: they could model hidden parameters in a system which we have no knowledge about. Therefore, not only ANN can improve the speed of PCF simulation, but also accuracy of the simulations. However, this final claim requires further analysis and experimentation.