Supplemental Document

Optics Letters

Structure-embedding network for predicting the transmission spectrum of a multilayer deep etched grating: supplement

PAN LIU,[1] YONGQIANG ZHAO,[1,*] SEONG G. KONG,[2] AND CHAOLONG TANG[1]

[1] *School of Automation, Northwestern Polytechnical University, Xi'an 710072, China*
[2] *Department of Computer Engineering, Sejong University, Seoul 05006, Republic of Korea*
[*] *Corresponding author: zhaoyq@nwpu.edu.cn*

Supplement DOI: https://doi.org/10.6084/m9.figshare.21493833

Parent Article DOI: https://doi.org/10.1364/OL.476383

Supplement of Structure-embedding network for predicting the transmission spectrum of MDEG

Pan Liu[1], Yongqiang Zhao[1,*], Seong G. Kong[2], Chaolong Tang[1]

[1] *School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China*

[2]*Department of Computer Engineering, Sejong University, Seoul, 05006 Korea*

1. The details about Pearson correlation coefficient (PCC) metrics.

Pearson correlation coefficient is used to measure the correlation (linear correlation) between two variables X and Y, and its value is between -1 and 1, as shown in Eq. (3). Both are used to evaluate the prediction performance of models.

$$\overline{\hat{Y}_i} = \frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i \tag{1}$$

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n}Y_i \tag{2}$$

$$r = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{\hat{Y}})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(\hat{Y}_i - \overline{\hat{Y}})}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{3}$$

Where n represents the sum of the test set, $\hat{Y}_i$ denotes the predicted spectrum, Y_i is the ground-truth. The $\overline{\hat{Y}}$ and $\overline{Y}$ are the average spectrum of the predicted spectrums and the ground-truth spectrums, respectively.

2. Normalization of the network input

The structure and training details of DNN and SEmNet are the same. We also normalized each class parameter in every batch according to Eq. (4). The other settings remain the same as before for improving the prediction accuracy. Experiment results demonstrate that the prediction accuracy is higher without normalization, as in Table 1. Therefore, we set the network input as the initial structural parameters without normalization.

$$X_{std} = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{4}$$

Table 1. Average MSE loss on the test set with or without input normalization

Methods	DNN	Reduce Sampling points	Binary inputs	SEmNet
MSE with normalization	0.0038	0.0035	0.0018	0.0016
MSE without normalization	0.0025	0.0024	0.0014	0.0012

3. The effect of various encoding length on model prediction performance and verification of the learnable matrix K.

In order to explore the effect of different encoding lengths on the prediction performance of the model, we conducted experiments with various encoding lengths of 1350, 1400, 1450, 1550, and 1600 in sequence. Table 2 shows the MAE, MSE, and PCC metrics on the test set. The performance metrics did not change much with different encoding lengths.

Table. 2. The MAE, MSE, and PCC metrics on the test set with various encoding lengths

Loss	1350	1400	1450	1500	1550	1600
MAE	0.0126	0.0126	0.0124	0.0122	0.0125	0.0124
MSE	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012
PCC	0.9005	0.9007	0.9008	0.9028	0.9028	0.9030

Table. 3. The MAE, MSE, and PCC metrics on test set when input is one-hot vector.

Loss	MAE	MSE	PCC
One-hot vector as input	0.0179	0.0015	0.8661
Binary input	0.0178	0.0014	0.8747
SEmNet	0.0122	0.0012	0.9028

To validate the effectiveness of the learnable matrix K, we directly input the one-hot vector to the neural network. Table 3 demonstrates that the performance without matrix K is almost the same as that of the binary input method, which proves that the learnable matrix K is valid. The sparsity of one-hot vector will affect the prediction accuracy of the network. Regarding the one-hot vector directly as the input does not transfer this method to other works. When the parameter range is large or more parameters exist, the one-hot vector dimension will be very high, which may cause a dimension disaster.

4. Finding Optimal Size of the Learnable Matrix K

The encoding module increases the dimensionality of the structural parameters while the embedding module decreases it using a learnable matrix, which exploits an implicit relationship among the structural parameters. The output of the structure-embedding module contains detailed information and good for the prediction of DNN module. To optimize the size of K, we conducted multiple experiments with the value of *d* ranging from 16 to 60 in a step of 2 and kept other training details unchanged. An initial value of K is determined by the size of MDEG. Fig. 1 shows average mean-square errors of d on the test set. The optimal d value was 50. A suitable mismatch between input and output dimensionality of a neural network can improve the prediction accuracy.

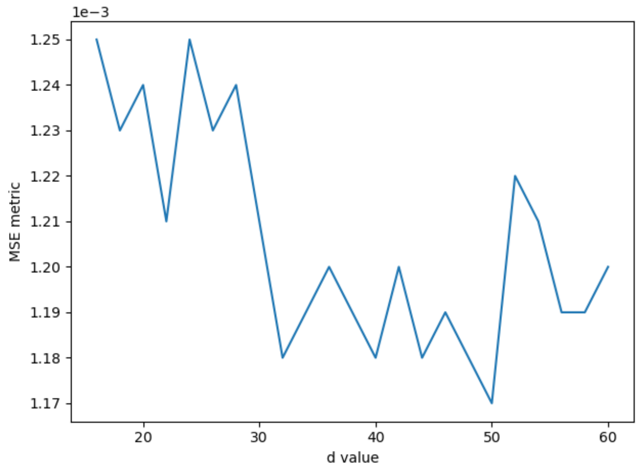

Fig. 1. Average MSE loss on the test set with the value of d ranging from 16 to 60 in steps of 2.

5. The hypothesis testing about the SEmNet and the binary input method

To illustrate the difference between Binary input method and SEmNet statistically, we perform the dataset random splitting and the corresponding training and testing separately for three times. These three experiments, as shown in Table 4-6, are denoted as realization_1, realization_2, and realization_3, which are used for hypothesis testing between the SEmNet and the binary input method.

Table 4. The average MSE loss on test set in three experiments

Methods	realization_1	realization_2	realization_3	Mean	Std($\times 10^{-5}$)
Binary Input	0.00141	0.00135	0.00136	0.00137	3.24
SEmNet	0.00117	0.00118	0.00119	0.00118	1

Table 5. The average MAE loss on test set in three experiments

Methods	realization_1	realization_2	realization_3	Mean	Std($\times 10^{-4}$)
Binary Input	0.0175	0.0181	0.0178	0.00178	3
SEmNet	0.0122	0.0123	0.0125	0.00123	1.58

Table 6. The average PCC metric on test set in three experiments

Methods	realization_1	realization_2	realization_3	Mean	Std($\times 10^{-3}$)
Binary Input	0.8727	0.8724	0.8791	0.8747	3.1
SEmNet	0.9040	0.9037	0.9007	0.9028	1.5

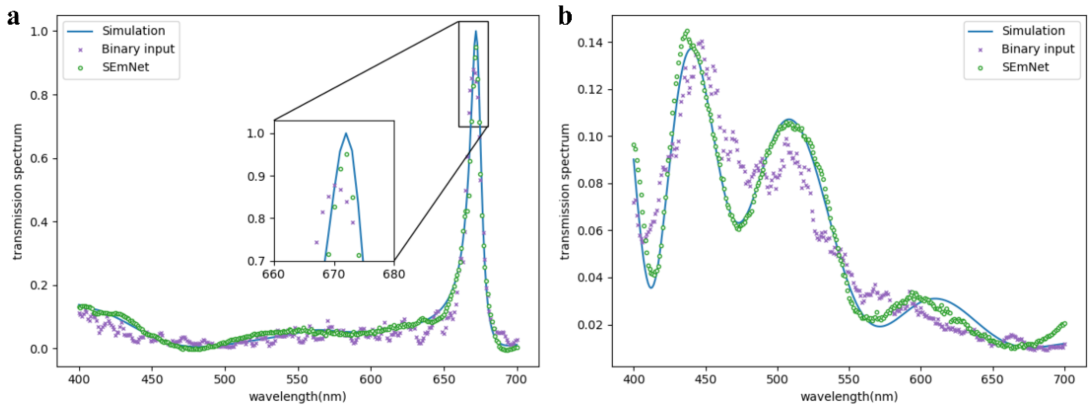

Fig. 2. The results of SEmNet and Binary input method. The green line represents the result of SEmNet and the purple line denotes the result of the Binary input method.

Fig. 2 shows the prediction performances of SEmNet and Binary input method in narrowband transmission spectrum and oscillation response. From the enlarged portion of the spectrum at the resonant frequency in Fig. 2(a), the prediction effect of the SEmNet at the peak is better than the binary input method. The dimension of embedding vector is adjusted by setting the hyperparameter d value, which alleviates the huge dimensionality mismatch of the neural network. This operation improves the prediction accuracy at the resonance, which is not achieved by binary input method. Fig. 2(b) also shows that the SEmNet is more accurate than the binary input method in predicting the complex response, such as oscillation response. The structure-embedding module exploits the relationship among structural parameters, which may not be available in the hard-coding method.

The t-test is mainly used for a normal distribution with a small sample size and an unknown population standard deviation. We assume that the metrics of the SEmNet and the binary input method obey the normal distribution. When the population variances are unknown and the two sample variances satisfy the following condition, as shown in Eq. (5), Independent two-sample t-test [1] is suitable for testing whether the means of two independent normal samples or similar normal samples are equal.

$$s_1^2 > 2s_2^2 \quad or \quad s_2^2 > 2s_1^2 \tag{5}$$

Where s_1^2 and s_2^2 represent the variances of two groups of samples. Let us define:

Null hypothesis H_0: The metrics on test set of the SEmNet and the binary input method are equal.

Alternative hypothesis H_1: The metrics on test set of the SEmNet and the binary input method

are unequal.
Use the Satterthwaite approximation[2] to construct statistics *t*:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{6}$$

The degree of freedom is given as Eq. (7):

$$v \approx \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{s_1^4}{n_1^2 v_1} + \frac{s_2^4}{n_2^2 v_2}} \tag{7}$$

$$v_1 = n_1 - 1 \tag{8}$$

$$v_2 = n_2 - 1 \tag{9}$$

Where $\overline{X_1}$, $\overline{X_2}$ are the mean of two groups of samples, n_1, n_2 are the size of two groups of samples, v_1, v_2 are the degrees of freedom of two groups of samples. The significant level α is set to 0.05. Then, $t_{\frac{\alpha}{2},v}$ represents the t-distribution value with a quantile of $\frac{\alpha}{2}$ and v degrees of freedom. The $t_{\frac{\alpha}{2},v}$ is obtained by inquiring the t-distribution table.

Table 7. The value of each variable in the hypothesis testing

Metric	v	t	$t_{\frac{\alpha}{2},v}$
MSE	2	-12.18	4.3027
MAE	3	-34.45	3.1824
PCC	3	14.17	3.1824

It can be obtained that $|t| > t_{\frac{\alpha}{2},v}$ from the table 7. Therefore, the H_0 is rejected. The performance on test set of the SEmNet and binary input is unequal. Generally, Hypothesis testing needs a large number of samples as support, which requires sufficient realizations of the training and testing experiments. A single realization takes 13 hours and we can not provide a lot of samples for hypothesis testing. From the table 4-7 and Fig 2, however, we have reason to believe the SEmNet outperforms the binary input method statistically.

6. Reference

(1) Welch, B. L. (1947). "The generalization of "Student's" problem when several different population variances are involved". Biometrika. 34 (1–2): 28–35.

(2) Spellman, Frank R. (12 November 2013). Handbook of mathematics and statistics for the environment. Whiting, Nancy E. Boca Raton. ISBN 978-1-4665-8638-3. OCLC 863225343.