OSNR Estimation Providing Self-Confidence Level as Auxiliary Output From Neural Networks

Takahito Tanimura, Takeshi Hoshida, *Member, IEEE*, Tomoyuki Kato, *Member, IEEE*, and Shigeki Watanabe, *Member, IEEE, Member, OSA*

(Invited Paper)

***Abstract*—Accurate optical monitors are critical for automating operations of fiber-optic networks. Deep neural network (DNN) based optical monitors have been investigated as accurate optical monitors to leverage a large amount of data obtained from fiber-optic networks. Although DNN-based optical monitors have been trained and tested to ensure the given accuracy criteria, this does not ensure sufficient accuracy under unexpected conditions, that is, out of test conditions, e.g., a newly developed modulation format that is not included in the test dataset. Thus, it is necessary to prepare a monitor to assess the current accuracy of a DNN-based optical monitor's output for robust automation of networks. We present a DNN-based optical monitor that simultaneously outputs an optical signal-to-noise ratio and its uncertainty information using a dropout method at the inference phase. This monitor was evaluated in cases in which the DNNs were trained with either a limited number of records or partially missing records in a training dataset. The proposed monitor successfully informed that own output has large uncertainties due to a limited amount of training data or a missing part in training dataset. Additionally, to improve an accuracy of estimated uncertainty, the number of partial neural networks by dropout at the inference phase was optimized. This is a valuable step toward designing robust "self-driving" optical networks.**

***Index Terms*—Coherent detection, deep neural networks, optical fiber communication, optical monitoring.**

I. Introduction

MACHINE-LEARNING-BASED optical monitoring [1], which can learn a mapping between the optical-fiber-channel parameters and the properties of the detected signal at the receiver, is vital to automate network operations and managements in future optical fiber communications [2], [3]. Figure 1 shows a schematic diagram of network automation based on such optical monitoring, which is able to perceive a real-time optical physical state and feed the information back to the network controller. On the network controller, a workflow enabling a decision-making based on the optical-monitor results is carried out, e.g., adjusting launch power, varying the adopted modulation format, and re-routing light-paths. Here the decision-making in the workflow relies on measured values from the optical monitor, such as the optical signal-to-noise ratio (OSNR).

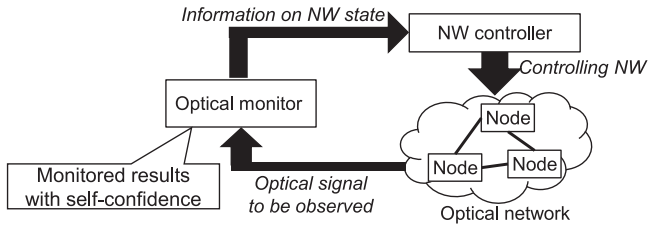

Fig. 1. High-level concept of optical network automation assisted by optical monitor providing self-confidence information. NW: network.

Considering practical situations, a measurement process in an optical monitor often produces an error that is difference between a monitored output and a true value. Thus, the automated workflow based on the monitor results may result in sub-optimal operation and/or an unexpected side-effect due to the error. To assess this risk, it is necessary to determine the uncertainty of the measured value from the optical monitor, which is the basis of the decision in the workflow. Workflows are automatically processed based on the monitored results without human intervention in normal cases. If the uncertainty of the monitor output is higher than a given threshold, the automated process should be stopped and all the monitored results should be passed to a human operator for better decision-making.

From the viewpoint of an optical monitor providing uncertainty information, there have been many investigations on shallow and deep neural network (DNN)-based optical monitors for accurate estimation but point-estimated values without model uncertainty. With various pre-process to extract features that are fed into a neural network, relatively simple neural networks have been proposed and investigated [4]–[10] to estimate various physical parameters such as the OSNR, chromatic dispersion (CD), polarization mode dispersion (PMD), and nonlinearity. To overcome manually provided pre-process to extract features, the use of more layers that act as automatic feature extractors by representation learning, i.e., using DNNs, has been investigated [11]–[17]. A DNN-based optical monitor that can skip manual feature engineering still only provides point estimations and cannot provide model uncertainty (or "confidence" in a model's output).

Manuscript received October 1, 2018; revised December 7, 2018 and January 20, 2019; accepted January 20, 2019. Date of publication February 6, 2019; date of current version April 2, 2019. *(Corresponding author: Takahito Tanimura.)*

The authors are with Fujitsu Laboratories Ltd., Kawasaki 211-8588, Japan (e-mail: tanimura.taka@jp.fujitsu.com; hoshida@jp.fujitsu.com; kato.tom@jp.fujitsu.com; s.watanabe@jp.fujitsu.com).

Digital Object Identifier 10.1109/JLT.2019.2895730

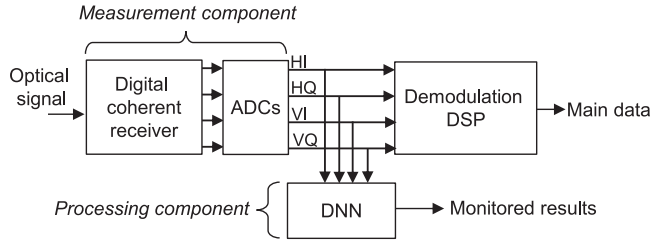

Fig. 2. Schematic diagram of deep neural network (DNN)-based optical monitor.

Besides neural networks, Gaussian process regression [18], [19] (GPR) is also used as another tool to enable optical monitoring. GPRs are mathematically grounded tools to determine model uncertainty. This is a powerful alternative to an optical monitor providing uncertainty information [20], [21]; however, there is an issue with GPRs in that there is computational complexity at the inference phase with a large amount of training data points. This is further discussed in Section II.C.

To overcome the above limitations, we previously presented an enhanced DNN-based OSNR monitor to output not only point-estimated OSNR but also predicted model uncertainty as auxiliary output [22]. This is enabled by random dropout of a trained DNN's neurons at the inference phase [23]. Unlike GPRs, the computation complexity of DNN with dropout at the inference phase does not depend on the number of training data points. Thus it is suitable to a situation characterized by an enormous number of available training data points.

In this study, we extended this work [22] by optimizing the number of partial DNNs generated in dropout at the inference phase to minimize the fluctuation in the predicted model uncertainties. We discuss the pros and cons of this extended DNN-based optical monitor by comparing DNN/dropout with a GPR. We also experimentally evaluated the model uncertainty predicted with this extended DNN-based monitor in two cases: a limited number of points and partially missing points in a training dataset. The evaluation results for the predicted model uncertainties correlated with actual uncertainties that were the standard deviation of point-estimated OSNR values.

The paper is organized as follows. In Section II, we presents our extended DNN-based optical monitor providing self-confidence as auxiliary output by using dropout at the DNN's inference phase. In Section III, we presents the experimental setup and DNN used in this study. After presenting the experimental results and accompanying discussion in Section IV, we conclude in Section V with a brief summary.

II. DNN-Based Optical Monitor With Uncertainty Information

A. Deep Neural Network-Based Optical Monitor

Figure 2 illustrates a schematic diagram of the DNN-based optical monitor [15]. A DNN-based optical monitor, which is composed of an off-the-shelf digital coherent receiver as a measurement component and a DNN as a processing component, is used to extract useful information for human network operators and/or network control programs.

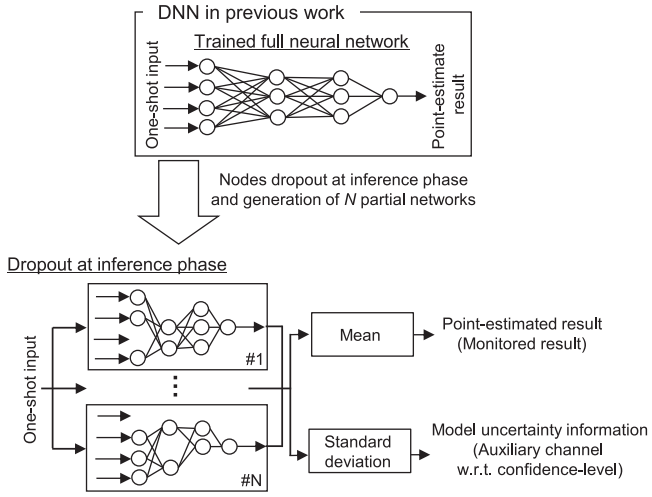

Fig. 3. Schematic diagrams of dropout at inference.

The digital coherent receiver measures an incoming optical signal and captures the waveform corresponding to the in-phase and quadrature phase of both horizontal and vertical polarization of the signal. The measured waveform data contain all information of the received optical field within the receiver bandwidth. Note that this measurement component can provide an enormous amount of data necessary for training DNNs thanks to the high-speed sampling rate (typically several dozen GSa/s) of analog-to-digital converters (ADCs) in the coherent receiver.

The processing component is used by the DNN that provides flexibility and versatility for processing. The DNN can program itself through its training phase and has the potential for versatile transformation to extract the value that corresponds to any waveform distortion embedded in the signal received from coherent reception. Thus, a DNN-based optical monitor has the potential to estimate many types of transmission impairments that are not only OSNR but also other physical impairments such as CD and PMD, which is unexpected in the design stage but needed later, from a measured dataset.

B. Auxiliary Output for Uncertainty Information

In our previous study [22], we enhanced the DNN-based optical monitor with auxiliary output that returns a result that provides the confidence level of the main monitored result. The key to developing such a monitor is using dropout in the DNN of the optical monitor. Standard dropout [24], [25] is widely used in training a DNN to avoid overfitting. Standard dropout samples a partial DNN from a full DNN by probabilistically deleting the full DNN's nodes with a probability of p for every input point and every forward and backward pass through the model training. At the inference phase, we use all nodes scaled by $1/(1-p)$. Standard dropout can improve the performance of a DNN but only provides point estimation results.

To predict model uncertainty as well as point estimation, dropout must be used at the inference phase, as shown in Fig. 3. This method for obtaining information on uncertainty from trained DNN models was proposed and developed by Gal

TABLE I
COMPARISON DNN/DROPOUT WITH GP

	GPR	DNN with dropout at inference
Uncertainty output	Available	Available
Overfitting	No concern	Of concern
To be configured for model instance	Kernel function	Number of layers, neurons, type of activation function
Computational complexity at inference phase	Depending on $N^{\dagger}$	Depending on $M^{\ddagger}$

†number of training data points ‡number of neurons

et al. [23]. Their method [23] for predicting model uncertainty performs an N-times forward pass through a DNN with dropout at the inference phase. Note that the scaling of $1/(1-p)$ is not used in this inference. The N results generated from N stochastic partial networks from dropout are averaged to predict the point-estimation results. To predict uncertainty, we collect the N results and take their standard deviation.

Using dropout at the inference phase can extract uncertainty information from existing trained DNNs without changing those DNNs. We introduce dropout at inference to DNN-based optical monitors so that they can provide both point-estimated monitoring results and their uncertainties.

C. Comparison of DNN/Dropout With GPR

In this section, we discuss and compare DNN with dropout at the inference phase described in the previous section with another method that is more established to treat uncertainty, i.e., a GPR [18], [19]. A standard DNN is a parametric model, which is a nonlinear mapping $\boldsymbol{y} = f(\boldsymbol{x}, \boldsymbol{w})$ from input $\boldsymbol{x}$ to output $\boldsymbol{y}$ that is governed by a weight vector $\boldsymbol{w}$ of adaptive parameters [24]. During the learning phase, a set of training data is used to obtain a point estimate of the parameter vector. Once the training is complete, the training data are discarded and predictions for new inputs are based on the learned $\boldsymbol{w}$. On the other hand, a GPR is a class of machine learning methods in which the training data points are kept and used also during the inference phase. A GPR is a non-parametric and probabilistic method of machine learning, which measures the similarity between training data points as the kernel function to predict the value for an unknown test point. Note that GPRs can naturally provide not only point-estimates but also a model's uncertainty information thanks to these probabilistic properties. This is a significant advantage of GPRs for our current objective, i.e., to obtain uncertainty information of model output.

Table I is a summary of the pros and cons of the two methods (GPR and DNN with dropout at the inference phase) to obtain uncertainty information of model output. Naturally, both methods support uncertainty output. Thanks to the Bayesian approach, a GPR is free from overfitting. On the other hand, we still have to be concerned with overfitting with DNNs, although it has been proved that simple learning algorithms such as stochastic gradient descent (SGD) [24] are guaranteed to find global minima on the training objective of DNNs in polynomial time according to latest research [26]. This result may relax our concern regarding overfitting on over-parameterized DNNs.

One of the possible issues with using a GPR is the need to construct valid kernel functions specialized for each specific task. For the application considered in this paper, valid kernel functions, which include the similarity of any two inputs that correspond to two of HI, HQ, VI, and VQ at different OSNRs, is not clear (note that H, V, I, and Q respectively represent horizontal and vertical polarizations and the in-phase and quadrature-phase components of an optical field). Thus it is difficult to find valid kernel functions specialized for this task. Although this is still a practicality issue, from a mathematical viewpoint, recent research [27] suggests a possible solution through emulating a DNN with any activation function (e.g., rectified linear unit (ReLU) [28]) by using a GPR with a specific kernel function. This research revealed that an equivalence between infinitely wide 'deep' neural networks and GPs, expanding well-known previous results of a single-layer neural network [19].

From a practical viewpoint, a major drawback of a GPR is that the required computational complexity at the inference phase depends on the number of training data points. This hinders the application of GPRs for many domains characterized by large training data sets. Traditionally in a GPR, it is likely difficult to calculate a large data set that contains over a few thousand data points even with various approximation techniques. On the other hand, the required computational complexity of DNNs at the inference phase does not depend on the number of the training data points, but mainly depends on the number of their neurons. In other words, GPRs are suitable for a case with a small number of data points; in contrast, DNNs have an advantage for cases with a large number of data points.

In our current application, i.e., optical monitoring at a digital coherent transceiver, available computational resource for inference a trained model should be restricted. Nevertheless, the number of measurable data points (i.e., the available number of training data points) might be enormous thanks to high-speed ADCs on the digital coherent transceiver. This may justify the use DNN/dropout to evaluate the uncertainty of model output in this specific case. Moreover, when we assumed that DNN-based optical monitors that were trained to output point-estimates have already been deployed over fiber-optic networks, we can reuse the existing DNN-based optical monitors for obtaining the uncertainty information without re-training of the DNN-based monitors. This enables the continuous update of a monitoring framework of fiber-optic networks without additional effort such as either replacement or re-training of monitors.

III. DEEP NEURAL NETWORK AND EXPERIMENTAL SETUP

A. Deep Neural Network Used in This Study

Figure 4 shows the DNN used in this study. The input for the DNN has four channelized electric fields in the digital coherent receiver: 512 samples $\times$ 4 channels to each HI, HQ, VI, and VQ sampled and digitized by the receiver. We convolved these time-series data over the time axis in the convolutional (conv.) layers, i.e., 1D convolution with trainable filter weights. We used multiple filters for these convolutional layers; thus, the number

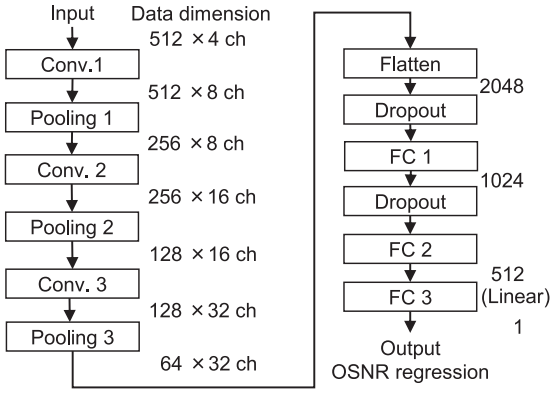

Fig. 4. DNN used in this study.

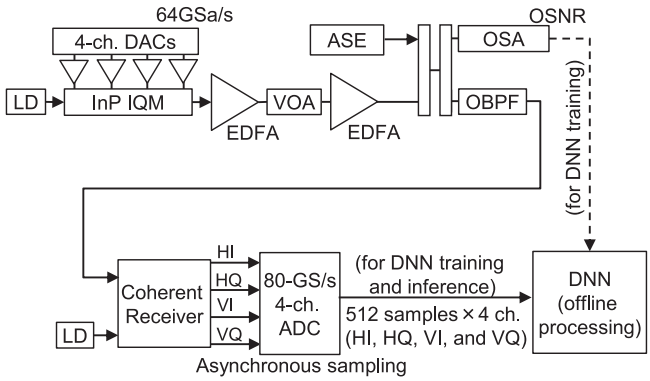

Fig. 5. Experimental and simulation setup. InP IQM: indium phosphide in-phase and quadrature-phase modulator, LD: laser diode, EDFA: erbium doped fiber amplifier, VOA: variable optical attenuator, ASE: amplified spontaneous emission source, OBPF: optical band-pass filter.

of output and input channels differed through each layer. All convolutional layers had the activation function of an ReLU [28]. The pooling layers used max pooling with strides of 2. These layers reduced the length of data by half, e.g., from 512 to 256. The data were flattened before the first fully connected (FC) layer, going from 64 samples × 32 channels to 2048 samples, as shown in Fig. 4. The flattened data were served to the FC layers. In FCs 1 and 2, dropout $(p = 0.5)$ was used at both training and inference phase. For output, linear regression was used in FC 3. A batch normalization technique [29] was also used to prevent overfitting.

The DNN was trained with supervised learning with a back-propagation and mini-batch SGD algorithm with a controlled learning rate by using the Adam optimizer [30] with the TensorFlow library [31]. The losses intended to be minimized were defined as the mean squared error.

B. Experimental Setup

Figure 5 shows the experimental setup. At the transmitter, an external cavity laser (∼25 kHz linewidth) was used as a light source for the channel at 193.3 THz. An Indium Phosphide DP-IQ-modulator was driven by the drive signals generated by a four-channel digital-to-analogue converter (DAC) with a sampling rate of 64 Gsample/sec and physical resolution of 8 bits. The DAC generated Nyquist-filtered (roll-off factor = 0.01) 16-Gbaud (GBd) dual-polarization quadrature phase shift keying (DP-QPSK) signals with pilot CW tones for carrier recovery (details of the transceiver digital signal processing (DSP) are given in a previous study [32]). The modulated signal was sent to an erbium-doped fiber amplifier, and additional amplified spontaneous emission noise was loaded to vary the received OSNR from 11 to 30 dB. The actual received OSNR was measured using an optical spectrum analyzer (OSA) and used as the desired DNN output in the training (this information is not necessary at the test/inference phase).

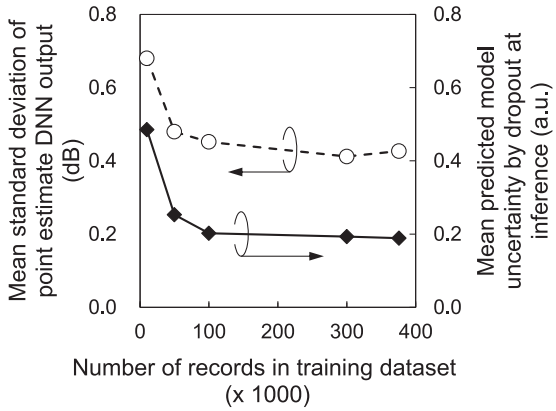

Fig. 6. Actual and predicted model uncertainty as a function of number of records in training datasets.

At the receiver, the local oscillator (∼25 kHz linewidth) was superimposed with the signal in a polarization-diversity optical 90° hybrid. The outputs of the hybrid were connected to four balanced photo-detectors. The resulting signals were digitized using four ADCs with a sample rate of 80 Gsample/sec. The digital samples were processed in an offline manner with a desktop computer equipped with 512-GByte memory, 2.2-GHz Intel Xeon E5 CPUs, and an Nvidia Tesla P100 GPU. Note that the DP-QPSK signals were modulated by different random bit sequences generated using the Mersenne Twister method with a different seed for each training and test/inference of the DNN to avoid an over-fitted evaluation. We confirmed correct reception of data by evaluating the bit error rate after demodulation DSP (not shown in this Fig. 5).

IV. Results and Discussion

First, we investigated the relationship between the actual standard deviation of point-estimated outputs and predicted model uncertainties. To generate different DNN models, DNNs were trained with datasets having different numbers of records: 10,000, 50,000, 100,000, 300,000, and 375,700. At the inference phase, we generated 200 partial networks to obtain one point-estimated OSNR and one piece of uncertainty information. The actual standard deviation of point-estimated DNN outputs, which was calculated using the 10,000 pieces of test data including different OSNRs from 11 to 30 dB, was increased while decreasing the number of records in the training datasets. Figure 6 shows predicted model uncertainties and the actual standard deviation of the point-estimated outputs of our extended DNN-based OSNR monitor.

The predicted model uncertainty correlated with the actual uncertainty that was the standard deviation. Thus, in possible

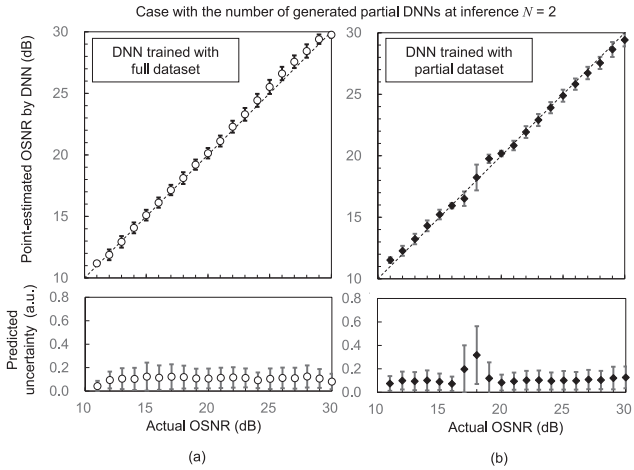

Fig. 7. Actual OSNRs vs. DNN-estimated OSNRs (upper) and predicted model uncertainties (lower) from DNN trained with (a) full and (b) partial dataset, $N = 2$.

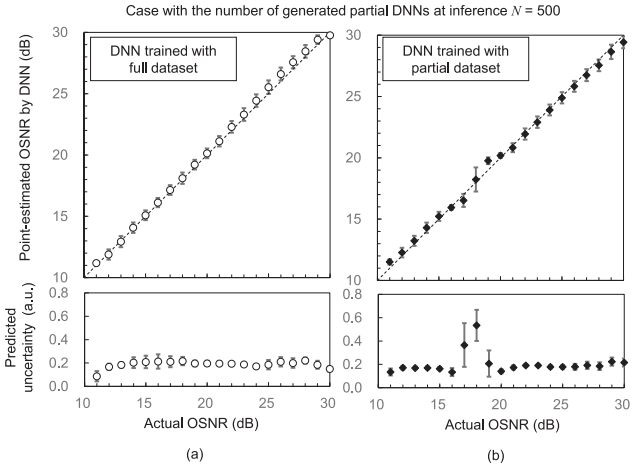

Fig. 8. Actual OSNRs vs. DNN-estimated OSNRs (upper) and predicted model uncertainties (lower) from DNN trained with (a) full and (b) partial dataset, $N = 500$.

use cases, the monitor could output high model uncertainty, i.e., a low confidence level in its output, to a network controller when the monitor output has large variance due to a limited amount of training data.

Next we investigated the model uncertainty with a lack of certain training data. To highlight the effect of this, we prepared artificial training datasets; data with OSNR of 17, 18, and 19 dB were deleted in the complete training dataset. To keep the total number the same, i.e., 375,700 records, the number of records of the other data (w.r.t. OSNR = 11–16 and 20–30 dB) was increased instead of that of the deleted data. After training the DNN by this "partial dataset" or the reference dataset having full data, we evaluated bias errors and standard deviations of the point-estimated outputs by using the trained DNNs.

To investigate the effect of *N,* i.e., the number of partial neural networks generated by the dropout at inference, we carried out inference of the trained DNNs with $N = 2$. The upper graphs in Figs. 7(a) and (b) show DNN-estimated OSNRs (i.e., point-estimated results) as a function of actual OSNRs for full and partial datasets, respectively. The whiskers on the point-estimated OSNRs show the standard deviation over multiple point-estimated OSNRs for different incoming signals. A lack of training data led to an increase in the standard deviation of DNN-estimated OSNRs (see points at OSNR of 17, 18, and 19 dB in the upper-right graph).

The lower graphs in Figs. 7(a) and (b) show predicted uncertainties (i.e., auxiliary channel outputs) from each DNN trained with the full or the partial dataset, respectively. The dropout at the inference phase successfully predicted the deviations of estimated OSNRs, as shown in the lower-right graph. The mean values of predicted uncertainties increased from 17 to 19 dB due to a lack of corresponding training data. The whiskers on the predicted uncertainties show the standard deviation over multiple outputs of the auxiliary channel (predicted uncertainties). Even with $N = 2$, the predicted uncertainty correlated with the actual standard deviation; however, the observed values of the predicted model uncertainty had a large variance. This large variance may introduce a risk that misleads decision-making in the field.

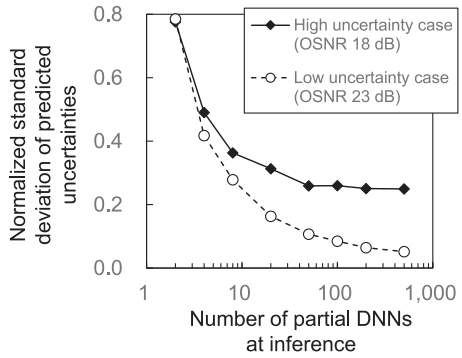

Fig. 9. Normalized standard deviation of predicted uncertainties as a function of number of partial DNNs at inference.

To mitigate this risk through reducing the fluctuation in predicted uncertainty by a DNN, we evaluated the fluctuation dependency against *N*. Figures 8(a) and (b) show DNN-estimated OSNRs as a function of actual OSNRs and show predicted uncertainties by each DNN trained with the full and partial dataset, respectively. We conducted inference of the trained DNNs with $N = 500$. While the mean values of predicted uncertainties also increased from an OSNR of 17 to 19 dB due to a lack of corresponding training data, it was observed to reduce the whiskers of predicted uncertainties (i.e., reducing the fluctuation in predicted model uncertainties) on both the DNNs trained with the full and partial datasets.

An increased *N* means that more randomly generated partial DNNs are used to predict uncertainty at inference. This suggests that we (possibly) use a more diverse ensemble of partial neural networks to estimate both OSNR and its uncertainty. This may be the reason we can reduce the fluctuation in predicted uncertainties with $N = 500$, compared with $N = 2$.

For detailed investigation of the fluctuation reduction, we focused on two cases: (1) high uncertainty corresponding to an OSNR of 18 dB and (2) low uncertainty corresponding to an OSNR of 23 dB in the DNN trained with the partial dataset. Figure 9 shows the standard deviation of predicted uncertainties normalized by the mean value of predicted uncertainties for

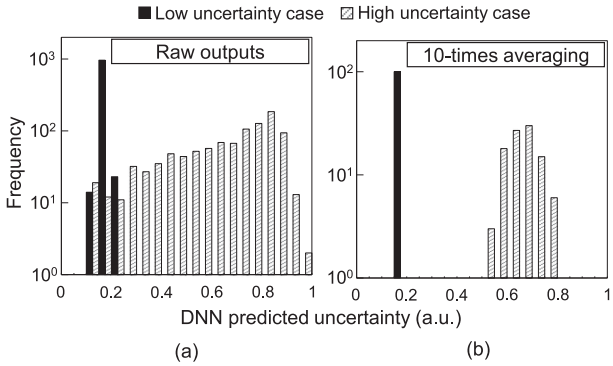

Fig. 10. Histograms of predicted uncertainties at high uncertainty (OSNR: 18 dB) and low uncertainty case (OSNR: 23 dB) for (a) row auxiliary channel output and (b) after 10-times averaging.

each OSNR. This graph shows that the fluctuation in predicted uncertainty decreased with an increase in N, and the reduction might be saturated around $N = 200$ for the high uncertainty case.

For a more detailed investigation, we analyzed the distribution of the predicted model uncertainty at an OSNR of 18 dB (high uncertainty case) and 23 dB (low uncertainty case). In this analysis, N was fixed at 200, as discussed in the previous paragraph. As shown in Fig. 10(a), we observed "long-tail" characteristics on the distribution of the predicted uncertainties for the high-uncertainty case. In other words, even by optimizing N, the observed values for predicted model uncertainty had a large variance. Unfortunately, there is still some risk in under/overestimating them. To solve this problem, we can simply delete this long-tail by averaging predicted uncertainties. Figure 10(b) shows a histogram of predicted uncertainties after 10-times averaging. The averaging mitigated the underestimation of predicted uncertainty and enabled the distinguishing of the high-uncertainty case from the low one.

V. Conclusion

We presented a deep neural network-based optical monitor that outputs both the optical signal-to-noise ratio and its confidence level at the same time by using dropout at the inference phase. The monitor was experimentally investigated with partially missing data and a limited number of training-data cases. By optimizing the number of partial neural networks at the inference phase and averaging, we distinguished the high-uncertainty case from the low one.

Acknowledgment

The authors would like to thank Ms. E. Katayama and Mr. K. Shiota of Fujitsu Kyushu Network Technologies Limited for their support in the data analysis.

References

[1] D. Zibar, M. Piels, R. Jones, and C. G. Schaeffer, "Machine learning techniques in optical communication," *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1442–1452, Mar. 2016.

[2] S. Yan *et al.*, "Field trial of machine-learning-assisted and SDN-based optical network planning with network-scale monitoring database," in *Proc. 43rd Eur. Conf. Opt. Commun.*, 2017, pp. 793–795.

[3] S. Oda *et al.*, "A learning living network with open ROADMs," *J. Lightw. Technol.*, vol. 35, no. 8, pp. 1350–1356, Apr. 2017.

[4] X. Wu, J. A. Jargon, R. A. Skoog, L. Paraschis, and A. E. Willner, "Applications of artificial neural networks in optical performance monitoring," *J. Lightw Technol.*, vol. 27, no. 16, pp. 3580–3589, Aug. 2009.

[5] J. A. Jargon, X. Wu, H. Y. Choi, Y. C. Chung, and A. E. Willner, "Optical performance monitoring of QPSK data channels by use of neural networks trained with parameters derived from asynchronous constellation diagrams," *Opt. Express*, vol. 18, no. 5, pp. 4931–4938, 2010.

[6] T. S. R. Shen, K. Meng, A. P. T. Lau, and Z. Y. Dong, "Optical performance monitoring using artificial neural network trained with asynchronous amplitude histograms," *IEEE Photon. Technol. Lett.*, vol. 22, no. 22, pp. 1665–1667, Nov. 2010.

[7] F. N. Khan, T. S. R. Shen, Y. Zhou, A. P. T. Lau, and C. Lu, "Optical performance monitoring using artificial neural networks trained with empirical moments of asynchronously sampled signal amplitudes," *IEEE Photon. Technol. Lett.*, vol. 24, no. 12, pp. 982–984, Jun. 2012.

[8] T. B. Anderson, A. Kowalczyk, K. Clarke, S. D. Dods, D. Hewitt, and J. C. Li, "Multi impairment monitoring for optical networks," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3729–3736, Aug. 2009.

[9] A. S. Kashi *et al.*, "Nonlinear signal-to-noise ratio estimation in coherent optical fiber transmission systems using artificial neural networks," *J. Lightw. Technol.*, vol. 36, no. 23, pp. 5424–5431, Dec. 2018.

[10] F. J. Vaquero Caballero *et al.*, "Machine learning based linear and nonlinear noise estimation," *J. Opt. Commun. Netw.*, vol. 10, pp. D42–D51, 2018.

[11] T. Tanimura, T. Hoshida, J. C. Rasmussen, M. Suzuki, and H. Morikawa, "OSNR monitoring by deep neural networks trained with asynchronously sampled data," in *Proc. Optoelectron. Commun. Conf./Int. Conf. Photon. Switching*, 2016, Paper TuB3-5.

[12] T. Tanimura *et al.*, "Deep learning based OSNR monitoring independent of modulation format, symbol rate and chromatic dispersion," in *Proc. Eur. Conf. Opt. Commun.*, 2016, Paper Tu.2.C.2.

[13] T. Tanimura, T. Hoshida, T. Kato, S. Watanabe, and H. Morikawa, "Data-analytics-based optical performance monitoring technique for optical transport networks (invited)," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2018, Paper Tu3E.3.

[14] T. Tanimura, T. Hoshida, T. Kato, S. Watanabe, and H. Morikawa, "Simple learning method to guarantee operational range of optical monitors," *J. Opt. Commun. Netw.*, vol. 10, pp. D63–D71, 2018.

[15] T. Tanimura, T. Hoshida, T. Kato, S. Watanabe, and H. Morikawa, "Convolutional neural network-based optical performance monitoring for optical transport networks," *J. Opt. Commun. Netw.*, vol. 11, pp. A52–A59, 2019.

[16] F. N. Khan *et al.*, "Joint OSNR monitoring and modulation format identification in digital coherent receivers using deep neural networks," *Opt. Express*, vol. 25, no. 15, pp. 17767–17776, 2017.

[17] D. Wang *et al.*, "Modulation format recognition and OSNR estimation using CNN-based deep learning," *IEEE Photon. Technol. Lett.*, vol. 29, no. 19, pp. 1667–1670, Oct. 2017.

[18] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (Adaptive Computation and Machine Learning). Cambridge, MA, USA: MIT Press, 2006.

[19] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[20] J. Wass, J. Thrane, M. Piels, R. Jones, and D. Zibar, "Gaussian process regression for WDM system performance prediction," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2017, Paper Tu3D.7.

[21] F. Meng *et al.*, "Field trial of Gaussian process learning of function-agnostic channel performance under uncertainty," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2018, Paper W4F.5.

[22] T. Tanimura, T. Kato, S. Watanabe, and T. Hoshida, "Deep neural network based optical monitor providing self-confidence as auxiliary output," in *Proc. Eur. Conf. Opt. Commun.*, 2018, Paper We1D.5.

[23] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. Conf. Mach. Learn.*, 2016, pp. 1651–1660.

[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[25] N. Srivastava *et al.*, "Dropout: A simple way to prevent neural networks from over fitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[26] Z. Allen-Zhu, Y. Li, and Z. Song, "A convergence theory for deep learning via over-parameterization," 2018, arXiv:1811.03962.

[27] J. Lee, J. Sohl-Dickstein, J. Pennington, R. Novak, S. Schoenholz, and Y. Bahri, "Deep neural networks as Gaussian processes," in *Proc. Int. Conf. Learn. Represent.*, 2018.

[28] X. Glorot, A. Bordes, and Y. Bengui, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
[29] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[30] J. Ba and D. Kingma, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, 2015.
[31] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Oper. Syst. Des. Implementation*, 2016, pp. 265–283.
[32] T. Tanimura, T. Hoshida, T. Kato, S. Watanabe, M. Suzuki, and H. Morikawa, "Throughput and latency programmable optical transceiver by using DSP and FEC control," *Opt. Express*, vol. 25, no. 10, pp. 10815–10827, May 2017.

Takahito Tanimura received the B.S. and M.S. degrees in physics from the Tokyo Institute of Technology (Tokyo Tech), Tokyo, Japan, in 2004 and 2006, respectively, and the Ph.D. degree in electrical engineering from The University of Tokyo, Tokyo, Japan, in 2018.

Since 2006, he has been with Fujitsu Laboratories Ltd., Kawasaki, Japan, where he has been engaged in the research and development of digital coherent optical communication systems. From 2011 to 2012, he was with the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany. His research interests include digital signal processing and machine learning for large-scale nonlinear systems.

Dr. Tanimura is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Physical Society of Japan (JPS) and is currently on the Editorial Committee of the *IEICE Transactions on Communications* (Japanese Edition) and the Technical Program Committee of the Optical Fiber Communication Conference.

Takeshi Hoshida (S'97–M'98) received the B.E., M.E., and Ph.D. degrees in electronic engineering from The University of Tokyo, Tokyo, Japan, in 1993, 1995, and 1998, respectively.

Since 1998, he has been with Fujitsu Laboratories Ltd., Kawasaki, Japan, where he has been engaged in the research and development of dense wavelength division multiplexing optical transmission systems. From 2000 to 2002, he was with Fujitsu Network Communications, Inc., Richardson, TX, USA. Since 2007, he has also been with Fujitsu Ltd., Kawasaki, Japan.

Dr. Hoshida is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Japan Society of Applied Physics (JSAP).

Tomoyuki Kato (S'05–M'06) was born in Saitama, Japan, in 1979. He received the B.E., M.E., and Dr.Eng. degrees in electrical engineering from Yokohama National University, Yokohama, Japan, in 2001, 2003, and 2006, respectively.

From 2006 to 2009, he was a Research Associate with the Precision and Intelligence Laboratories, Tokyo Institute of Technology. In 2009, he joined Fujitsu Laboratories Ltd., Kawasaki, Japan. His current research includes nonlinear optical signal processing.

Dr. Kato is a member of the IEEE Photonics Society and the Institute of Electronics, Information and Communication Engineers of Japan (IEICE).

Shigeki Watanabe (M'93) received the B.S. and M.S. degrees in physics from Tohoku University, Sendai, Japan, in 1978 and 1980, respectively, and the Ph.D. degree in electrical engineering from The University of Tokyo, Tokyo, Japan, in 1997.

In 1980, he joined Fujitsu Ltd., and since 1987, he has been with Fujitsu Laboratories Ltd., Kawasaki, Japan, where he is engaged in advanced phonic technologies in the field of optical communications. His current research interests include nonlinear optical signal processing and ultrafast photonics.

Dr. Watanabe is a member of the Optical Society of America (OSA), and the Institute of Electronics, Information and Communication Engineers of Japan (IEICE).