Mappatura speditiva tridimensionale e multi-temporale mediante UAV. I casi di Pescara del Tronto e Accumoli

This article has been accepted for publication in IEEE Journal of Selected Topics in Quantum Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTQE.2022.3151878

On Cooperative Fault Management in Multi-domain Optical Networks Using Hybrid Learning

Xiaoliang Chen, Che-Yu Liu, Roberto Proietti, Jie Yin, Zhaohui Li, S. J. Ben Yoo, *Fellow, IEEE, Fellow, OSA*

(Invited Paper)

***Abstract*—This paper presents a hybrid learning approach for cooperative fault management in multi-domain optical networks (MD-ONs). The proposed approach relies on a broker-based MD-ON architecture for coordination of inter-domain service provisioning. We first propose a self-supervised learning design for soft failure detection. The self-supervised learning design makes use of a clustering algorithm for extracting normal and abnormal patterns from optical performance monitoring data and a supervised learning-based classifier trained with the learned patterns for online detection. To facilitate high soft failure detection accuracy in the absence of sufficient abnormal data for training, the proposed design estimates model uncertainties during predictions and identifies instances associated with high uncertainties as also soft failures. Then, we extend the self-supervised learning design and present a federated learning framework for the broker plane and DMs to learn cooperatively while complying with the privacy constraints of each domain. Finally, a data-driven soft failure localization scheme that operates by analyzing the patterns of data is proposed as a complement to the existing approaches. Performance evaluations indicate that the self-supervised learning design can achieve soft failure detection accuracy of up to $\sim 97\%$ with $0.01\% - 0.04\%$ false alarm rate, while federated learning enables DMs to realize $> 90\%$ soft failure detection rates in the cases of highly unbalanced data distribution (two of the three domains possess zero abnormal data for training).**

***Index Terms*—Multi-domain optical networks (MD-ONs), soft failure detection and localization, self-supervised learning, model uncertainty, federated learning.**

I. Introduction

As the underlying infrastructure of the Internet, optical networks carry traffic generated from heterogeneous applications (e.g., social networking, multimedia) at the rate of a few hundred Gigabits up to Terabits per second per wavelength [1], [2]. A single component failure in optical networks can lead to severe service disruptions. Therefore, effective fault management schemes for optical networks are of vital importance.

X. Chen and Z. Li are with the Guangdong Provincial Key Labratory of Optoelectronic Information Processing Chips and Systems, Sun Yat-sen University, Guangzhou 510275, China. Z. Li is also with the Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), 519000, China (Email: xlichen@ieee.org, lzhh88@mail.sysu.edu.cn).

C. Liu is with the Department of Computer Science, University of California, Davis, Davis, CA 95616, USA (Email: cyliu@ucdavis.edu).

R. Proietti and S. J. B. Yoo are with the Department of Electrical and Computer Engineering, University of California, Davis, Davis, CA 95616, USA (Email: rproietti@ucdavis.edu, sbyoo@ucdavis.edu).

J. Yin is with the Fortinet Corporation, Sunnyval, CA, 94086, USA (Email: jyin22@ucsc.edu).

X. Chen and C. Liu are co-first authors.

Manuscript received Oct. 11, 2021.

Previous studies have reported extensive designs for failure detection and localization in optical networks targeting hard failures (e.g., fiber cuts) [3], [4]. Unlike hard failures that can bring down connections immediately, soft failures refer to incidents that cause moderate and gradual performance degradation, for instance, device aging, equipment malfunctioning, misconfigurations, and physical-layer attacks [5], [6]. Prompt detection and localization of soft failures is highly desired as it enables optical networks to operate with lower margins (thus, higher resource efficiency) and prevents expensive hard failures that these soft failures may eventually evolve to. However, effective management of soft failures is not straightforward because they are often covert. Traditional approaches typically apply threshold-based policies on particular network parameters, such as the monitored signal power of a lightpath, which suffer from poor flexibility. In particular, determining a proper value of threshold entails network experts investigating the behaviors of the related parameters under a specific system setup. As network conditions may change (due to changes of topology, deployment of new equipment or services, etc.), such approaches need to be constantly refined, prohibiting the fast evolution of optical networks. Besides, some soft failures can exhibit sophisticated patterns that cannot be easily characterized by a simple threshold [7].

Lately, machine learning (ML) has emerged as one of the key enabling techniques for building next-generation optical networks. ML equips network control and management (NC&M) systems with the potential to learn network rules or operation policies automatically from data (network traces, past experiences, etc.), thus, largely enhancing the intelligence of network operations [8]–[10]. In this context, researchers have proposed a number of ML-based approaches for cognitive fault management in optical networks [7], [11]–[19]. In [7], the authors proposed finite state machine-based algorithms for detecting and identifying bit-error-rate (BER) degradation caused by four types of soft failures, namely, signal overlap, tight filtering, gradual drift, and cyclic drift. Later in [11], the same authors studied ML approaches for localizing these soft failures. In [12], Shahkarami et al. conducted a performance comparison between different ML algorithms when applied to soft failure detection and identification in optical networks. The authors of [13] presented a neural network-based classifier design for detecting abnormal signal power variations under different failure modes. In [14], the authors developed two classification algorithms, aiming at detecting and identifying jamming signal attacks of various intensities in optical networks. In [17], the authors demonstrated a neural network model taking as input the power spectrum density of

This article has been accepted for publication in IEEE Journal of Selected Topics in Quantum Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTQE.2022.3151878

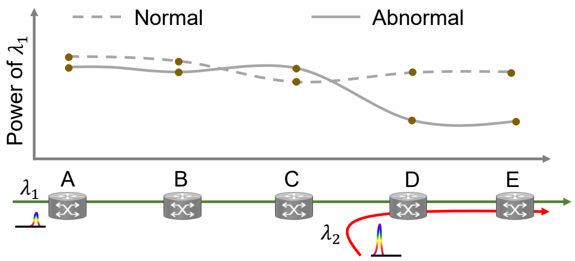

Fig. 1. An example of soft failure caused by jamming signal attack.

a received signal and the filter tap coefficients to locate the malfunctioning optical switch.

All of the above works apply supervised learning approaches, which demand for large sets of labeled data for successful training. A major factor that limits the applicability of these approaches is the scarcity of abnormal data since real optical networks operate under normal states most of the time. Moreover, the trained supervised learning models can only recognize soft failures having been identified by human experts, i.e., related features and labels must be provided in the training sets. To overcome this issue, our previous work in [15] devised a hybrid unsupervised and supervised learning approach working directly on unlabeled data. The rationale is that abnormal behaviors typically show unique patterns deviating from those of normal ones [20], and therefore, by analyzing the patterns of data through data clustering (unsupervised learning), arbitrary type of soft failures can potentially be detected. Then, a supervised learning-based classifier is trained with the patterns learned by unsupervised learning to facilitate online detection with low time complexity. Fig. 1 shows an example of soft failure caused by a jamming signal attack launched at node D [6]. By examining the power of λ_1 along the routing path, we can observe a pattern that differs from a normal fluctuation (profiled by the dashed line) despite that the power values may still locate within normal ranges. Following a similar idea, the authors of [16] proposed a dual-stage approach, which makes use of only BER and signal power information at the first stage while exploiting a more comprehensive digital spectrum features at the second stage if a soft failure is detected. More recent works also made attempts to detect soft failures by learning a mapping from original data to a space where normal and abnormal data can be more easily distinguished [18], [19].

Nevertheless, existing works only considered single-domain scenarios where network administrators possess full domain visibility. It is known that the Internet infrastructure is composed of multiple autonomous systems/domains and assuring high quality and availability of inter-domain services is indispensable [21]. Based on domain privacy considerations, domain managers (DM) tend to advertise only limited intra-domain information, making effective fault management in multi-domain optical networks (MD-ONs) a non-trivial task.

In this work, we propose to realize cooperative fault management in MD-ONs with a hybrid ML approach. The proposed design takes advantage of a broker-based MD-ON architecture for coordination of inter-domain service provisioning. We first refine our previous work in [15] to present a self-supervised learning approach for soft failure detection. Our approach incorporates estimations of model uncertainty during inferences to facilitate high soft failure detection rate even in the absence of sufficient abnormal data for training. Then, a federated learning framework is proposed to enable cooperative learning between the broker plane and DMs while complying with the domain privacy constraint. Finally, we present a data-driven soft failure localization scheme that operates by analyzing the patterns of data. Performance evaluations conduced with data collected using the VPItransmissionMaker™ Optical Systems simulator verify the effectiveness of the proposed design.

The rest of the paper is organized as follows. In Section II, we provide an overview of the cooperative fault management architecture. In Sections III and IV, we detail the hybrid learning design and show the corresponding performance evaluations. Finally, Section V summaries the key paper contributions.

II. Network Architecture

Optimizing service provisioning in MD-ONs entails a powerful NC&M system that can well coordinate the operations of multiple domains while complying with the domain autonomy constraints [21]. Thanks to the unprecedented network programmability offered by software-defined networking (SDN) and the recent breakthroughs in ML, previous works have demonstrated a broker-based MD-ON architecture facilitating cognitive inter-domain networking [8], [22]. This work exploits such an architecture and develops a cooperative fault management framework.

Fig. 2 illustrates the layout of the proposed framework. We consider an MD-ON of three hierarchies, namely, data, DM, and broker planes. The data plane carries aggregated client traffic with lightpaths established by proper configuration of optical transmission equipment, i.e., transponders, wavelength-selective switches, etc. Each DM adopts the SDN paradigm to operate its data plane. Specifically, a centralized SDN controller is employed to interact with the data plane equipment for collecting network state information and distributing configuration instructions. For instance, making use of network telemetry techniques [23], the SDN controller of DM-2 can actively surveil the status (e.g., signal power, noise level, channel utilization) of lightpaths *LP-A* and *LP-B* that traverse its domain. Above the SDN controller, each DM deploys various NC&M and service provisioning applications [24]–[27], such as fault management and routing and spectrum assignment, for generating operation policies for the related tasks. When ML is incorporated in the design of these applications, observe-analyze-act cycle-based cognitive networking can be realized: *i)* observe network state, *ii)* perform data analytics, *iii)* take actions leveraging the knowledge acquired. Lying in the top hierarchy, the newly introduced broker plane coordinates inter-domain service provisioning operations. It works with DMs according to mutual service level agreements (SLAs) rather than a dictate-comply principle. In particular, based on the specifications defined in the SLAs, DMs can report different degrees of abstracted domain data [e.g., several available

This article has been accepted for publication in IEEE Journal of Selected Topics in Quantum Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTQE.2022.3151878

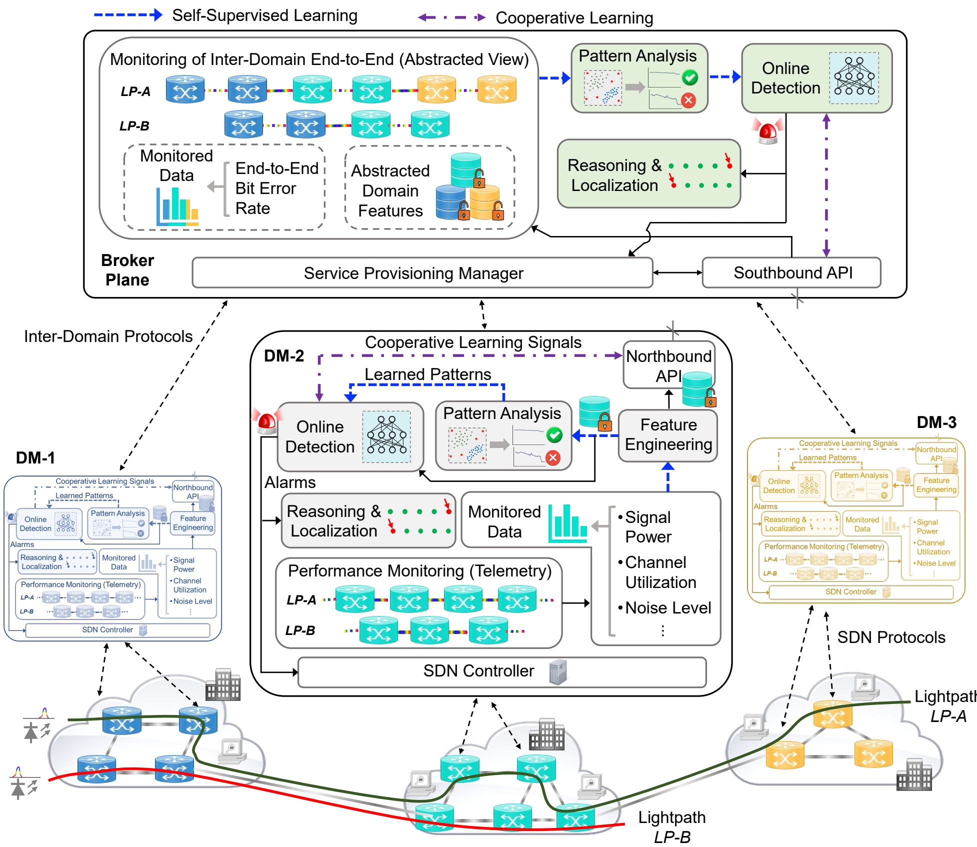

Fig. 2. Cooperative fault management framework in broker-based MD-ONs. DM: domain manager.

wavelengths on a virtual link (intra-domain path segment) between two border nodes] to the broker plane, which in turn, recommends the service schemes to be used. This way, domain confidentiality can be preserved.

Benefiting from the broker-DM synergy, we devise a hybrid learning approach for fault management, particularly, soft failure detection and localization, in MD-ONs. Within each domain, the fault management application (FMA) first preprocesses and extracts relevant features from the raw optical performance monitoring (OPM) data with the feature engineering module. Then, the FMA performs pattern analysis on the obtained features using unsupervised learning (i.e., clustering) approaches. Because abnormal network states exhibit patterns dissimilar to those of normal observations and only occur occasionally, data clustering enables to detect soft failures by identifying outlying instances that cannot form clusters. Hence, applying unsupervised learning eliminates the need for prior knowledge about abnormal network behaviors and potentially allows to detect unseen soft failures. On the other hand, such an approach can suffer from scalability issues as it requires revisiting the whole data set every time a new received instance is to be processed. In this context, we introduce a self-supervised learning mechanism that further trains a supervised learning model (e.g., a neural network-based classifier) with the patterns learned by unsupervised learning. Once trained, the supervised learning model can perform online soft failure detection, while its complexity only relates to the model attributes (e.g., scale). Therefore, scalability concerns can be mitigated. Upon detecting a soft failure, the FMA triggers the soft failure localization and reasoning functionalities and meanwhile raises an alarm to alert the SDN controller. The broker plane employs an FMA similar to that of a DM for fault management of inter-domain services. The broker plane FMA can work with the received abstracted domain features by itself or initiate cooperative learning procedures with domain FMAs when the possessed data are inadequate for certain tasks or when DMs are willing to share knowledge for improved performance. As for the latter case, the broker plane and domain-level supervised learning models exchange cooperative learning signals, for instance, gradients in a federated learning scheme, which will be detailed in the next section. In the case when a soft failure associated with an inter-domain service is detected and localized by the broker plane, it informs related DMs to conduct further inspections and works with them to reconfigure the service if necessary.

III. Hybrid Learning Design

This section elaborates on the hybrid learning approach, including the self-supervised learning design for soft failure detection, the application of federated learning for broker-DM synergy, and a data-driven soft failure localization design.

A. Soft Failure Detection by Self-supervised Learning

Unsupervised Learning: We apply the density-based clustering algorithm (dubbed DBSCAN) proposed in [28]. The rationale is that DBSCAN is able to detect clusters of arbitrary

This article has been accepted for publication in IEEE Journal of Selected Topics in Quantum Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTQE.2022.3151878

(nonspherical) shapes and does not require the number of clusters to be specified as in other clustering algorithms (e.g., K-means). DBSCAN involves three key elements: distance metric $dist(\cdot)$, ε, and $MinPts$. Distance metric (e.g., Euclidean distance) evaluates the similarity between data instances. ε sets the distance threshold for two instances to be regarded as neighbors, while $MinPts$ defines the minimum number of neighboring nodes for an instance to be counted as a core node. Algorithm 1 shows the procedures of DBSCAN. The idea is to iteratively form clusters by repeating the following steps: *i)* starting with a random and unvisited core node (*Lines 4-5*), and *ii)* continually expanding the cluster until all the neighbors of the core nodes in the cluster have been included (*Lines 6-11*). Finally, instances that cannot be clustered are detected as outliers/soft failures (*Line 15*). More details about the procedures of DBSCAN can be found in [28].

Algorithm 1: Procedures of DBSCAN.

Input: Data set $\mathcal{S}$, ε, $MinPts$
Output: Cluster set $\mathcal{C}$, outlier set $\mathcal{U}$
1 initialize $\mathcal{C}$ as an empty set;
2 calculate $dist(x, x'), \forall x, x' \in \mathcal{S}$;
3 **for** *each* $x \in \mathcal{S}$ **do**
4   **if** *x is an unvisited core node* **then**
5   set x as the first node of the new cluster;
6   store the neighboring nodes of x in Λ;
7   **while** *Λ is not empty* **do**
8   expand the new cluster with Λ;
9   overwrite Λ with the neighboring nodes of the instances in Λ;
10   remove from Λ the instances having been clustered;
11   **end**
12   **end**
13   store the new cluster in $\mathcal{C}$;
14 **end**
15 store the remaining instances in $\mathcal{U}$;

Supervised Learning: We design the supervised learning module with a neural network-based classifier that predicts whether each data instance is abnormal or not. Note that, the distribution of OPM data can be highly biased, i.e., the vast majority of data represent normal network states. Lacking abnormal data for training can result in poor soft failure detection accuracy when a regular neural network model is employed. Unlike regular neural networks whose weights θ are deterministic variables, a Bayesian neural network (BNN) [29] models the posterior distribution of θ given an observation of data $\mathcal{S}$, i.e., $p(\theta|\mathcal{S})$. Based on the Bayes rule, we have,

$$p(\theta|\mathcal{S}) = \frac{p(\mathcal{S}|\theta)p(\theta)}{\int p(\mathcal{S}|\theta)p(\theta)d\theta}, \tag{1}$$

where $p(\theta)$ is the prior belief about the distribution of θ and $p(\mathcal{S}|\theta)$ is the conditional distribution of $\mathcal{S}$ given θ. BNNs allow us to measure model uncertainty in predictions, which can assist in meeting the aforementioned challenge. More specifically, BNNs facilitate detecting abnormal data instances that are rarely seen in the training sets as they typically correspond to higher model uncertainties compared with normal instances.

Inferring $p(\theta|\mathcal{S})$ with Eq. 1 is often intractable because we need to enumerate all possible θ. A practical implementation of BNNs is Monte-Carlo (MC) dropout [30]. To realize MC dropout, we simply introduce dropout layers to the neural network-based classifier to enable deactivation of a certain portion of neurons (with a probability, e.g., 0.1) during both training and inference phases. The prediction for each instance x can be obtained by performing a large number of MC simulations (each time a neural network model f(θ^i) is sampled) and estimating the expectation of inference, i.e.,

$$\begin{aligned} &p(y=c) = \frac{1}{N}\sum_{i\in[1,N]} p(y=c|x;\theta^i), \\ &y = \arg\max_{c\in\mathcal{C}} \left\{p(y=c)\right\}, \end{aligned} \tag{2}$$

where $\mathcal{C}$ is the set of classes. In the meantime, we can evaluate the model uncertainty by calculating the mutual information [30] by,

$$\begin{aligned} I = &-\sum_{c\in\mathcal{C}} p(y=c)\log p(y=c) + \\ &\frac{1}{N}\sum_{c\in\mathcal{C}}\sum_{i\in[1,N]} p(y=c|x;\theta^i)\log p(y=c|x;\theta^i). \end{aligned} \tag{3}$$

Finally, the proposed approach detects an instance x as abnormal if the prediction result y corresponds to the abnormal class or the measured uncertainty is higher than a threshold U_{th}. We will describe the method to decide a proper setting for U_{th} in Section IV. Note that, applying the MC dropout technique for estimating classification uncertainties would introduce additional computational overheads. Such overheads can be mitigated by parallelizing the feed-forward calculations during each inference. Alternatively, one can switch to a different uncertainty estimation technique that does not rely on sampling of models, for instance, by training a deterministic neural network that directly learns the distributions of class probabilities [31]. The application of such techniques in soft failure detection will be left as one of our future studies.

B. Broker-DM Synergy by Federated Learning

To realize cooperative learning between the broker plane and DMs without the need for data sharing, we extend the design discussed in Section III-A by applying a federated learning mechanism. Table I summarizes the procedures of the proposed design. Each of the broker plane and DMs employs a neural network classifier of the same architecture. Let θ_G and $\theta_d (d \in \mathcal{D})$ denote the sets of model weights of the broker plane and domain d, respectively. The learning process starts with the broker plane randomly initializing θ_G and consists of K training iterations (*Steps* 2-4). In each iteration, DMs first synchronizes their models with θ_G copied from the broker plane (*Step* 2). In *Step 3*, each DM performs independent training of M epochs on the local data set $\mathcal{S}_d$ using a standard

This article has been accepted for publication in IEEE Journal of Selected Topics in Quantum Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTQE.2022.3151878

TABLE I
PROCEDURES OF THE FEDERATED LEARNING DESIGN.

Step 1:	the broker plane randomly initializes θ_G.
Step 2:	each DM downloads θ_G and assigns $\theta_G \rightarrow \theta_d$.
Step 3:	each DM trains its classifer with $\mathcal{S}_d$ and submits the encrypted gradients $\triangledown_{\theta_d}$ to the broker plane.
Step 4:	the broker plane aggregates the received gradients and update θ_G accordingly.
Step 5:	execute *Steps 2-4* for K times.

training algorithm (e.g., Adam [32]) at learning rate η_d. Then, DMs submit the encrypted gradients $\triangledown_{\theta_d}$ to the broker plane. Finally, in *Step 4*, the broker plane aggregates the received gradients by,

$$\triangledown_G = \frac{1}{\sum_{d\in\mathcal{D}} |\mathcal{S}_d|} \sum_{d\in\mathcal{D}} |\mathcal{S}_d| \triangledown_{\theta_d}, \quad (4)$$

where $|\mathcal{S}_d|$ represents the number of data instances in $\mathcal{S}_d$, and updates θ_G with learning rate η_G.

C. Data-driven Soft Failure Localization

Traditional correlation-based failure localization schemes require the routing information from multiple lightpaths [33], which can be difficult to obtain in MD-ONs. While recent studies have investigated several ML-assisted cognitive approaches [34], these approaches rely on large sets of data related to different failure scenarios for training, leading to restricted practicability. In this work, we exploit the results from data clustering and propose a data-driven soft failures localization approach as a complement to the existing designs. The idea of the proposed approach is to look into the patterns of data and attempt to localize positions where deviations of pattern originate. For the sake of clarity, we reuse $\mathcal{S}$ to denote the set of data composed of OPM information from different locations of lightpaths, for instance, readings of signal power at domain border nodes for inter-domain lightpaths. In other words, $\mathcal{S}$ convey spatial characteristics of lightpaths. For each abnormal instance x detected, we pick an instance $s^* \in \mathcal{S}$ that has the minimum distance to x. If s^* is an abnormal instance, we presume that x falls into the pattern same as that of s. Otherwise, we calculate the distance between x and s^* in each dimension, which can be represented as,

$$h(x, s^*) = [h^1(x, s^*), \cdots, h^J(x, s^*)]. \quad (5)$$

where J is the number of dimensions of each data instance. We also obtain the gradient of $h(x, s^*)$ as,

$$\begin{aligned} \triangledown h(x, s^*) =& [h^1(x, s^*), h^2(x, s^*) - h^1(x, s^*), \cdots, \\ & h^J(x, s^*) - h^{J-1}(x, s^*)]. \end{aligned} \quad (6)$$

Then, we infer the location where the pattern of x deviates from that of s^* [i.e., the location of soft failure, which can be a node (for intra-domain cases) or a domain (for inter-domain cases)] by computing,

$$j^* = \arg\max_j \left\{ h^j(x, s^*) + \triangledown^j h(x, s^*) \right\}. \quad (7)$$

TABLE II
LIST OF SYSTEM PARAMETER CONFIGURATIONS.

Modulation Format	Symbol Rate (Gbauds)	Launch Power (mWatt)	
		Co-Tx ID	Range
4 QAM	$23-28$	1	$1.3-3.6$
		2	$2.2-2.5$
		3	$0.22-0.25$
QPSK	$23-28$	1	$1.4-3.3$
		2	$2.2-2.4$
		3	$0.22-0.24$
8 PSK	$9-10$	1	$0.8-1.2$
		2	$1.5-1.6$
		3	$1.5-1.6$

Recall the example in Fig. 1, it is obvious that we get $j^* = 4$ and thus, successfully localize the attack at node D. For an inter-domain case, the broker plane informs the related DM to conduct further localization operations.

IV. PERFORMANCE EVALUATION

We valuated the performance of the proposed design with data collected using the VPItransmissionMaker™ Optical Systems simulator. Fig. 3 shows the system setup with the simulator for data generation. We set up lightpaths consisting of five nodes (from node A to E). The three I/Q modulator blocks, each fed by eight wavelength division multiplexing (WDM) lasers, were used to generate signals of interest (λ_2) as well as to inject background signals. Each Co-Tx operated at 224 Gbaud and adopted one of {4QAM, QPSK, 8PSK} as the modulation format[1]. Each node was connected to its neighboring nodes by standard single-mode fibers of 100 km. We compensated the fiber loss between two nodes by using gain-controlled amplifiers with a noise figure of 4 dB. The optical spectrum analyzers placed along the lightpaths monitored the power of signals continuously.

We emulated various network configurations by setting up lightpaths with different: *i)* modulation formats, *ii)* symbol rates, and *iii)* launch power. Table II summarizes the configuration of the system parameters. To emulate evolving network conditions, we created time-varying link loads (by changing the number of background signals inserted) and meanwhile introduced a 0.3 dB or 0.6 dB reduced gain to the amplifier in node C. We picked 15 out of these configurations to further create abnormal network states. Specifically, we followed a common practice in literature and applied one of the five soft failure cases to each of the configurations: Case 1) amplifier malfunctioning [12], [16] (emulated by introducing 10 dB attenuation to the amplifier in node D), Case 2) high-power jamming attacks [6], [14] (emulated by injecting a signal of

[1]Despite that 4QAM and QPSK are identical in their most known forms, the simulator modulates symbols with different sets of amplitudes and phases for the two modulation formats, leading to slightly different transmission characteristics.

This article has been accepted for publication in IEEE Journal of Selected Topics in Quantum Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTQE.2022.3151878

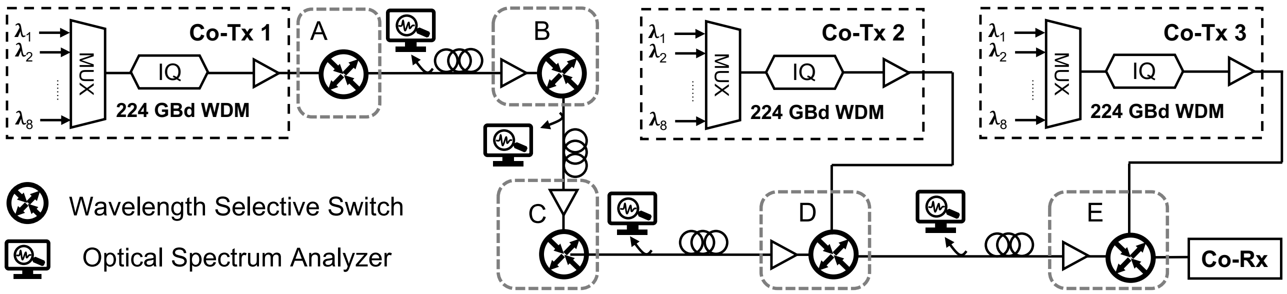

Fig. 3. System setup for data generation.

3 mWatt on λ_3 in node D), Case 3) misconfiguration [7] (emulated by activating simultaneously the laser working on λ_2 in Co-Tx 2), Case 4) laser central frequency drift [7], [16] [emulated by drifting the laser central frequency over a distance (half of channel spacing)], and Case 5) tight filtering [7], [16] (emulated by narrowing the bandwidth of the lowpass filter at the receiver side by $0.2\times$).

We collected the signal power monitored at each node and the BER values measured at the receiver side, and processed these information to generate two data sets, namely, temporal-characteristic and spatial-characteristic data sets. Each instance of the temporal-characteristic data set is composed of the BER values of a lightpath monitored at five continuous time points, describing the temporal behaviors of the lightpath. The spatial-characteristic data set conveys the spatial behaviors of lightpaths, i.e., each instance is a concatenation of the power monitored along a lightpath and the BER value at a specific system time. Overall, 3000 and 2880 normal instances were generated for two data sets, respectively, while both contain 15 abnormal instances.

A. Soft Failure Detection

1) Self-supervised Learning: We first assessed the performance of the self-supervised learning approach in soft failure detection, assuming that the whole data sets are possessed by a single entity (the broker plane or a DM). For data clustering, we implemented the DBSCAN algorithm with the Euclidean distance metric. We fixed $MinPts = 4$ and decided the setup for ε with the approach in [15] other than by trial and error that is typically used in training supervised learning models. Specifically, based on the distribution of the measured distances for a given data set, we set a range of ε and evaluate the evolution of the ratio of outliers detected as a function of ε. The dash-dot curves in Fig. 4 show the related results for the two data sets. When ε takes a relatively small value, most of the instances are detected as outliers as they can hardly be identified as neighbors and thereby form clusters. As ε keeps increasing, the ratios of outliers first decrease sharply and then turn stable from certain points. These inflection points reflect the correct setups of ε which allow normal instances in the majority to be clustered while leaving the abnormal instances outlied. That is, the reflection points indicate the marginal distances between normal instances. Such an approach is data-driven and can be executed automatically without human intervention. For the two data sets, we determined the setup

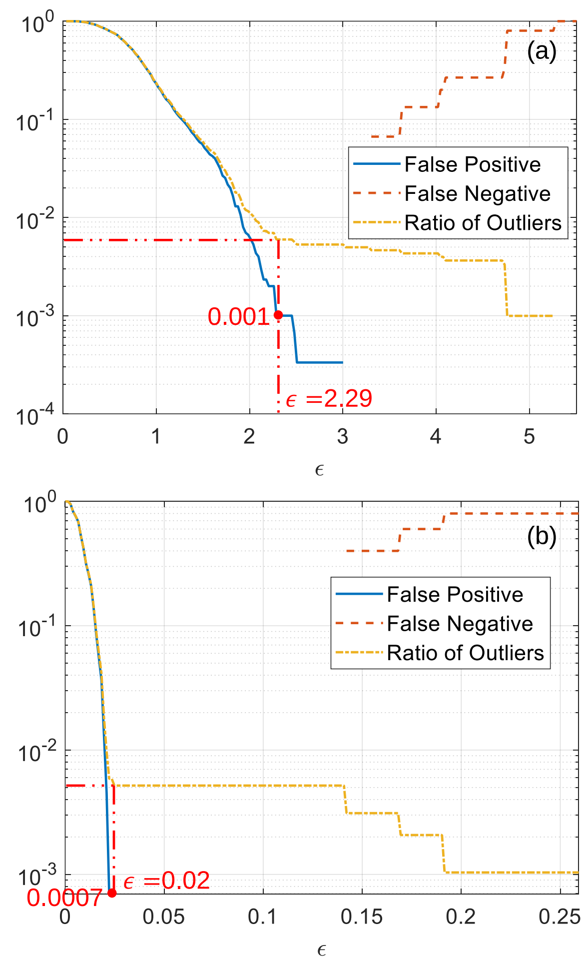

Fig. 4. Results of false positive rate, false negative rate, and ratio of outliers detected as functions of ε with (a) the temporal-characteristic and (b) the spatial-characteristic data sets.

of ε to be 2.29 and 0.02, respectively. We also plot the results of false positive and false negative rates in Fig. 4. It can be seen that with the setup of ε determined, the algorithm can achieve 100% soft failure detection rate (zero false negative rate) with negligible false positive (false alarm) rates, i.e., 0.1% and 0.07%. Note that, the results are presented in the logarithmic scale and therefore, break offs of curve represents zero values.

Based on the results from data clustering, we labeled the data sets and trained classifiers implemented by neural networks (NNs) of four layers, i.e., $[5, 10, 10, 2]$ and $[6, 10, 10, 2]$

This article has been accepted for publication in IEEE Journal of Selected Topics in Quantum Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTQE.2022.3151878

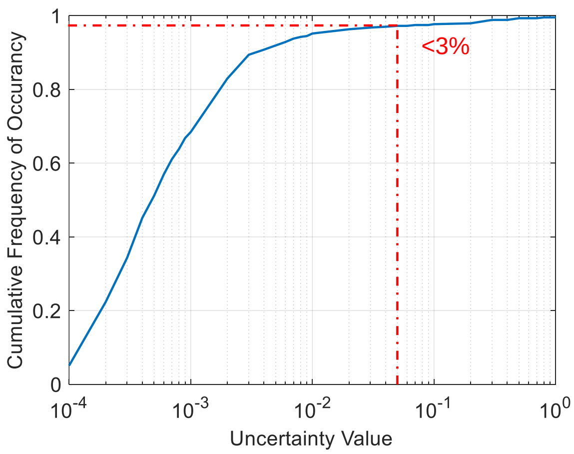

Fig. 5. Distribution of uncertainty for the temporal-characteristic data set.

for the two data sets, respectively. The hidden layers make use of *ELU* as the activation function while the output layers employ the *Softmax* function. For the proposed uncertainty-aided approach, we introduced a dropout rate of 0.1 for the hidden layers and conducted 100 feed-forward calculations to obtain the prediction for each instance (after the NNs had been trained). Fig. 5 shows the distribution of uncertainty for the temporal-characteristic data set when performing inference on a testing set (20% out of the entire data set). There exists an obvious trade-off in the choice of the uncertainty threshold U_{th} as a lower threshold facilitates soft failure detection but can lead to excessive false alarms. Similar to the method used for deciding the setup of ε, we can set U_{th} a value from which the uncertainty distribution curve flattens so that only a small number of inferences are with uncertainties higher than U_{th} (e.g., less that 3% in Fig. 5). In particular, we chose $U_{th} = 0.05$, which also applies to the spatial-characteristic data set. We compared the proposed approach with a baseline relying on purely the predictions from regular NNs without dropout. Table III summarizes the results of false negative and false positive rates for the temporal-characteristic data set with different proportions (denoted by γ) of data used for training. For $\gamma = 0.2$, 0.4, 0.6 and 0.8, the numbers of abnormal instances used for training and testing are $(3, 12)$, $(6, 9)$, $(9, 6)$, and $(12, 3)$, respectively. Each result is obtained by averaging the outcomes of 100 independent experiments. We can see that the proposed approach can achieve almost 90% soft failure detection rate when only 20% of the data are used for training, whereas the accuracy from the baseline is only $\sim 64\%$ in this setting. Such a performance gain only leads to a 0.3% raise in false positive rate. With 80% of data used for training, the false negative rate from the proposed approach decreases to 3.5%, while that from its counterpart is 15.7%. The resultant raise in false positive rate is still as low as 0.4%. Table IV shows the results for the spatial-characteristic data set, where similar observations can be drawn. With only 20% of data used for training, the proposed approach realizes a soft failure detection rate of 94% while introducing an addition of 0.01% to the false positive rate.

2) Federated Learning: Having demonstrated the effectiveness of the self-supervised learning approach, we next evaluated the benefit of cooperative learning between the broker plane and DMs in a multi-domain setting (an MD-ON of three domains). We considered two data division schemes, namely, <u>uniform</u> and <u>nonuniform</u>. In the uniform scheme, we evenly distributed the normal and abnormal data instances to the three DMs. Whereas in the nonuniform scheme, we created a biased distribution of abnormal data by assigning all the abnormal instances to DM-1. We compared the federated learning approach with an independent learning mechanism, where each DM trains its models independently with the local data sets. The federated learning models were implemented by the 'TensorFlow Federated' package, with $M = 10$ and $K = 30$. The uncertainty threshold was set to be 0.005 and 0.05 for federated learning and independently learning, respectively, according to the method discussed in the previous section. Again, we performed 100 experiments for each approach and data division scheme and obtained the averaged results. In all the experiments, 80% of the normal data and 60% of the abnormal data (i.e., nine instances) were used for training, and the rest were used for evaluation. Table V shows the results for the temporal-characteristic data set. Under uniform data distribution when DM possess independently and identically distributed (*i.i.d.*) data, federated learning does not dominate its counterpart. Federated learning leads to $3.5\% - 6\%$ higher false negative rates but also more than $10\times$ lower false positive rates, which is mainly caused by the different choices of the uncertainty threshold. However, the advantage of federated learning becomes evident when nonuniform data distribution was applied. Without abnormal data used for training, DM-2 and DM-3 can hardly detect soft failures (33.5% and 36.5%, respectively). By exploiting knowledge from data of multiple domains, federated learning enables DMs to achieve false negative and false positive rates comparable to those under uniform data distribution. Table VI shows the results for the spatial-characteristic data set. We can observe similar trends for the two approaches. Overall, the results verify the benefit

TABLE III
COMPARISON BETWEEN THE UNCERTAINTY-AIDED APPROACH AND THE BASELINE FOR THE TEMPORAL-CHARACTERISTIC DATA SET.

γ		0.2	0.4	0.6	0.8
f_n	Regular NN	0.359	0.260	0.195	0.157
	Uncertainty-Aided	0.101	0.067	0.047	0.035
f_p	Regular NN	0	0	0	0
	Uncertainty-Aided	0.003	0.003	0.003	0.004

TABLE IV
COMPARISON BETWEEN THE UNCERTAINTY-AIDED APPROACH AND THE BASELINE FOR THE SPATIAL-CHARACTERISTIC DATA SET.

γ		0.2	0.4	0.6	0.8
f_n	Regular NN	0.190	0.137	0.095	0.072
	Uncertainty-Aided	0.060	0.042	0.029	0.023
f_p	Regular NN	0	0	0	0
	Uncertainty-Aided	0.0001	0.0002	0.0001	0.0001

TABLE V
COMPARISON BETWEEN FEDERATED LEARNING AND INDEPENDENT LEARNING FOR THE TEMPORAL-CHARACTERISTIC DATA SET.

		Uniform Data Distribution			Nonuniform Data Distribution		
Domain ID		1	2	3	1	2	3
f_n	Independent Learning	0.055	0.045	0.020	0.005	0.665	0.635
	Federated Learning	0.090	0.105	0.070	0.070	0.090	0.100
f_p	Independent Learning	0.044	0.036	0.049	0.066	0.003	0.002
	Federated Learning	0.003	0.003	0.004	0.0004	0.0006	0.0005

TABLE VI
COMPARISON BETWEEN FEDERATED LEARNING AND INDEPENDENT LEARNING FOR THE SPATIAL-CHARACTERISTIC DATA SET.

		Uniform Data Distribution			Nonuniform Data Distribution		
Domain ID		1	2	3	1	2	3
f_n	Independent Learning	0	0	0	0	0.145	0.10
	Federated Learning	0.015	0.010	0.015	0	0.005	0
f_p	Independent Learning	0.058	0.053	0.067	0.049	0.002	0.001
	Federated Learning	0.0001	0.0001	0	0	0.0001	0.0001

TABLE VII
LOCALIZATION ACCURACY FOR EACH SOFT FAILURE CASE.

Case ID	1	2	3	4	5
Accuracy	3/3	3/3	2/3	1/3	0/3

of cooperative learning in MD-ONs.

B. Soft Failure Localization

Finally, we assessed the performance of the proposed data-driven soft failure localization approach using the spatial-characteristic data set. The accuracy results are presented in Table VII. We can see that the proposed approach can successfully localize amplifier malfunctioning (Case 1) and high-power jamming attack (Case 2) as they lead to variations of signal power which can be captured by analyzing the patterns of data. Two of the three misconfigurations (Case 3) are correctly localized, as overlapping of signals also influences signal power but less notably compared with the first two cases. Whereas for Case 4 and Case 5, the proposed approach can hardly localize the positions where soft failures are introduced. The reason is that frequency drift and tight filtering have negligible impact on signal power [7], necessitating data sets representing more features than just signal power. We will leave this as one of our future works.

V. CONCLUSION

In this paper, we demonstrated a hybrid learning approach for cooperative fault management in MD-ONs. We first presented a self-supervised learning design for soft failure detection. The self-supervised learning design makes use of a clustering algorithm for extracting normal and abnormal patterns from data and a supervised learning-based classifier aided by model uncertainty analysis for online detection. Then, we proposed a federated learning framework for cooperative learning between the broker plane and DMs. Finally, a data-driven soft failure localization scheme was presented. Performance evaluations show that the proposed self-supervised learning design can achieve high soft failure detection accuracy when only a few abnormal data instances are used for training and that federated learning enables effective knowledge sharing between DMs under highly unbalanced data distributions. A potential future research work could be investigating more comprehensive soft failure localization approaches exploiting the latest advances in ML, such as graph neural networks.

ACKNOWLEDGMENTS

This work was supported in part by the Key-Area Research and Development Program of Guangdong Province under Grant 2018B010114002, NSF award # ICE-T:RC 1836921, NSFC Project U2001601, and NSFC Project 62035018.

REFERENCES

[1] O. Gerstel, M. Jinno, A. Lord, and S. J. B. Yoo, "Elastic optical networking: a new dawn for the optical layer?" *IEEE Commun. Mag.*, vol. 50, pp. S12–S20, Apr. 2012.

[2] P. Lu, L. Zhang, X. Liu, J. Yao, and Z. Zhu, "Highly-efficient data migration and backup for big data applications in elastic optical inter-datacenter networks," *IEEE Netw.*, vol. 29, pp. 36–42, Sept./Oct. 2015.

[3] S. Ahuja, S. Ramasubramanian, and M. Krunz, "Single-link failure detection in all-optical networks using monitoring cycles and paths," *IEEE/ACM Trans. Netw.*, vol. 17, no. 4, pp. 1080–1093, 2009.

[4] M. Khair, B. Kantarci, J. Zheng, and H. T. Mouftah, "Optimization for fault localization in all-optical networks," *J. Lightw. Techn.*, vol. 27, no. 21, pp. 4832–4840, 2009.

[5] C. Mas, I. Tomkos, and O. Tonguz, "Failure location algorithm for transparent optical networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 8, pp. 1508–1519, Aug. 2005.

[6] S. Nina, M. Furdek, S. Zsigmond,, and L. Wosinska, "Physical-layer security in evolving optical networks," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 110–117, Aug. 2016.

[7] A. Vela, M. Ruiz, F. Fresi, N. Sambo, F. Cugini, G. Meloni, L. Potí, L. Velasco, and P. Castoldi, "BER degradation detection and failure identification in elastic optical networks," *J. Lightw. Technol.*, vol. 35, no. 21, pp. 4595–4604, Nov. 2017.

[8] X. Chen, R. Proietti, H. Lu, A. Castro, and S. J. B. Yoo, "Knowledge-based autonomous service provisioning in multi-domain elastic optical networks," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 152–158, Aug. 2018.

[9] F. Musumeci, C. Rottondi, A. Nag, T. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Commun. Surv. Tutor.*, vol. 21, no. 2, pp. 1383–1408, Secondquarter 2019.

[10] F. Khan, Q. Fan, C. Lu, and A. Lau, "An optical communication's perspective on machine learning and its applications," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 493–516, Jan. 2019.

[11] A. Vela, B. Shariati, M. Ruiz, F. Cugini, A. Castro, H. Lu, R. Proietti, J. Comellas, P. Castoldi, S. J. B. Yoo, and L. Velasco, "Soft failure localization during commissioning testing and lightpath operation," *J. Opt. Commun. Netw.*, vol. 10, no. 1, pp. A27–A36, Jan. 2018.

[12] S. Shahkarami, F. Musumeci, F. Cugini, and M. Tornatore, "Machine-learning-based soft-failure detection and identification in optical networks," in *Proc. Conf. Opt. Fiber Commun.*, Mar. 2018, paper M3A.5.

[13] D. Rafique, T. Szyrkowiec, H. Grießer, A. Autenrieth, and J. Elbers, "Cognitive assurance architecture for optical network fault management," *J. Lightw. Technol.*, vol. 36, no. 7, pp. 1443–1450, Apr. 2018.

[14] C. Natalino, M. Schiano, A. Giglio, L. Wosinska, and M. Furdek, "Field demonstration of machine-learning-aided detection and identification of jamming attacks in optical networks," in *Proc. Eur. Conf. Opt. Commun.*, Sept. 2018, pp. 1–3.

[15] X. Chen, B. Li, R. Proietti, Z. Zhu, and S. J. B. Yoo, "Self-taught anomaly detection with hybrid unsupervised/supervised machine learning in optical networks," *J. Lightw. Technol.*, vol. 37, no. 7, pp. 1742–1749, April 2019.

[16] L. Shu, Z. Yu, Z. Wan, J. Zhang, S. Hu, and K. Xu, "Dual-stage soft failure detection and identification for low-margin elastic optical network by exploiting digital spectrum information," *J. Lightw. Technol.*, vol. 38, no. 9, pp. 2669–2679, 2020.

[17] H. Lun, X. Liu, M. Cai, M. Fu, Y. Wu, L. Yi, W. Hu, and Q. Zhuge, "Anomaly localization in optical transmissions based on receiver DSP and artificial neural network," in *Proc. Conf. Opt. Fiber Commun.*, 2020, paper W1K.4.

[18] S. Liu, D. Wang, C. Zhang, L. Wang, and M. Zhang, "Semi-supervised anomaly detection with imbalanced data for failure detection in optical networks," in *Proc. Conf. Opt. Fiber Commun.*, 2021, paper Th1A.24.

[19] H. Lun, X. Liu, M. Cai, Y. Wu, M. Fu, L. Yi, W. Hu, and Q. Zhuge, "GAN based soft failure detection and identification for long-haul coherent transmission systems," in *Proc. Conf. Opt. Fiber Commun.*, 2021, paper Th4J.2.

[20] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection for discrete sequences: A survey," *IEEE Trans. Knowl. Data Engi.*, vol. 24, no. 5, pp. 823–839, May 2012.

[21] S. J. B. Yoo, "Multi-domain cognitive optical software defined networks with market-driven brokers," in *Proc. Eur. Conf. Opt. Commun.*, Sept. 2014, paper WE.2.6.3.

[22] G. Liu, K. Zhang, X. Chen, H. Lu, J. Guo, J. Yin, R. Proietti, Z. Zhu, and S. J. B. Yoo, "The first testbed demonstration of cognitive end-to-end optical service provisioning with hierarchical learning across multiple autonomous systems," in *Proc. Conf. Opt. Fiber Commun.*, 2018, paper Th4D.7.

[23] F. Paolucci, A. Sgambelluri, F. Cugini, and P. Castoldi, "Network telemetry streaming services in SDN-based disaggregated optical networks," *J. Lightw. Technol.*, vol. 36, no. 15, pp. 3142–3149, 2018.

[24] Z. Zhu, W. Lu, L. Zhang, and N. Ansari, "Dynamic service provisioning in elastic optical networks with hybrid single-/multi-path routing," *J. Lightw. Technol.*, vol. 31, no. 1, pp. 15–22, Jan. 2013.

[25] L. Gong, X. Zhou, X. Liu, W. Zhao, W. Lu, and Z. Zhu, "Efficient resource allocation for all-optical multicasting over spectrum-sliced elastic optical networks," *J. Opt. Commun. Netw.*, vol. 5, pp. 836–847, Aug. 2013.

[26] Y. Yin, H. Zhang, M. Zhang, M. Xia, Z. Zhu, S. Dahlfort, and S. J. B. Yoo, "Spectral and spatial 2D fragmentation-aware routing and spectrum assignment algorithms in elastic optical networks," *J. Opt. Commun. Netw.*, vol. 5, no. 10, pp. A100–A106, Oct 2013.

[27] L. Gong and Z. Zhu, "Virtual optical network embedding (VONE) over elastic optical networks," *J. Lightw. Technol.*, vol. 32, pp. 450–460, Feb. 2014.

[28] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. Int. Conf. Knowl. Discovery Data Mining*, Aug. 1996, pp. 226–231.

[29] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural networks," 2015, arXiv:1505.05424.

[30] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," 2016, arXiv:1506.02142.

[31] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2018, p. 3183–3193.

[32] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, arXiv:1412.6980.

[33] T. Panayiotou, S. P. Chatzis, and G. Ellinas, "Leveraging statistical machine learning to address failure localization in optical networks," *J. Opt. Commun. Netw.*, vol. 10, no. 3, pp. 162–173, Mar. 2018.

[34] H. Lun, X. Liu, M. Cai, M. Fu, Y. Wu, L. Yi, W. Hu, and Q. Zhuge, "Anomaly localization in optical transmissions based on receiver DSP and artificial neural network," in *Proc. Conf. Opt. Fiber Commun.*, 2020, paper W1K.4.

Xiaoliang Chen received the Ph.D. degree from the University of Science and Technology of China in 2016. He is an Associate Professor at the Sun Yat-Sen University. Prior to joining the Sun Yat-sen University, he was a postdoctoral researcher at the University of California, Davis (UC Davis). His research interests include optical networks, optical interconnection architectures for data centers, software-defined networking, and cognitive networking.

Che-Yu Liu was born in Tainan, Taiwan, in 1989. He received the B.S. degree in electrical and computer engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2012, and the M.S. degree in electrical and computer engineering, UC Davis, in 2016. He is currently working toward the Ph.D. degree in computer science, UC Davis. His research interest focuses on optical networks, software-defined networking, and machine learning.

Roberto Proietti received the M.S. degree in telecommunications engineering from the University of Pisa, Pisa, Italy, in 2004, and the Ph.D. degree in electrical engineering from Scuola Superiore Sant Anna, Pisa, in 2009. He is a Project Scientist with the Next Generation Networking Systems Laboratory at UC Davis. His research interests include optical switching technologies and architectures for supercomputing and data center applications, high-spectral-efficiency coherent transmission systems, and elastic optical networking.

Jie Yin received the M.S. degree from the University of California, Santa Cruz, in 2020. His research interests include optical networks and software-defined networking.

Zhaohui Li received the B.S. degree from the Department of Physics and M.S. degree from the Institute of Modern Optics, Nankai University, Tianjin, China in 1999 and 2002, respectively, and the Ph.D. degree from the Nanyang Technological University, Singapore, in 2007. He is working with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China, as a Professor, in 2009. His research interests are optical communication systems, optical signal processing technology and ultrafine measurement systems.

This article has been accepted for publication in IEEE Journal of Selected Topics in Quantum Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTQE.2022.3151878

S. J. Ben Yoo is a Distinguished Professor at UC Davis. His research at UC Davis includes 2D/3D photonic integration for future computing, cognitive networks, communication, imaging, and navigation systems, micro/nano systems integration, and the future Internet. Prior to joining UC Davis in 1999, he was a Senior Research Scientist at Bellcore, leading technical efforts in integrated photonics, optical networking, and systems integration. His research activities at Bellcore included the next-generation Internet, reconfigurable multiwavelength optical networks (MONET), wavelength interchanging cross connects, wavelength converters, vertical-cavity lasers, and high-speed modulators. He led the MONET testbed experimentation efforts, and participated in ATD/MONET systems integration and a number of standardization activities. Prior to joining Bellcore in 1991, he conducted research on nonlinear optical processes in quantum wells, a four-wave-mixing study of relaxation mechanisms in dye molecules, and ultrafast diffusion-driven photodetectors at Stanford University (BS 84', MS 86', PhD 91', Stanford University). Prof. Yoo is Fellow of IEEE, OSA, NIAC and a recipient of the DARPA Award for Sustained Excellence (1997), the Bellcore CEO Award (1998), the Mid-Career Research Faculty Award (2004 UC Davis), and the Senior Research Faculty Award (2011 UC Davis).