WDM System Stimulated Raman Scattering Spectrum and Tilt Prediction using CNN-based Transfer Learning

Shuang Xie[1], Rishu Raj[1], Dmitrii Briantcev[1], Zehao Wang[2], Tingjun Chen[2], Dan Kilper[1]
[1]CONNECT Centre, Trinity College Dublin, Ireland; [2]Duke University, Durham, NC, USA
Email: {xiesh, rajr, briantcd, dan.kilper}@tcd.ie, {zehao.w, tingjun.chen}@duke.edu

***Abstract*—Advanced optical network technologies, proposed to fulfil the projected demands of future communication networks, are riddled with numerous challenges. Conventional methods for addressing these issues often face scalability limitations in disaggregated systems, prompting a growing reliance on data-driven approaches. Nevertheless, accurately modeling physical phenomena, such as stimulated Raman scattering (SRS), remains difficult since conventional physics-based models can falter due to variations in transmission spans. Machine learning-based models are promising in this regard, but they require huge datasets for enhanced versatility. Another alternative approach is transfer learning (TL), which minimizes the need for extensive datasets by utilizing pre-trained models that incorporate relevant domain knowledge. In this paper, we design and evaluate transfer learning models based on convolutional neural networks (CNN). We use these models to predict the Raman tilt spectra across four fiber configurations in two different testbeds. We achieve less than 0.162 dB mean absolute error which is largely limited by the measurement accuracy.**

***Index Terms*—Transfer learning, CNN, Raman tilt, Optical Communication**

I. Introduction

The high bandwidth and low latency demands of next-generation wireless networks (e.g., 5G/6G) and wide-area networks (e.g., data center interconnects, DCI) depend heavily on optical infrastructure [1]. This infrastructure employs wavelength-division multiplexing (WDM) technology to enhance data transmission capacity [2]. However, as WDM signals propagate through the optical network, they encounter wavelength-dependent effects, primarily influenced by the erbium-doped fiber amplifier (EDFA) gain profile and fiber non-linearities [3]. One significant nonlinearity is stimulated Raman scattering (SRS), which transfers optical power from shorter wavelength signals to longer wavelength signals [4]. This interaction creates an asymmetric power distribution across the WDM spectrum, resulting in a spectral tilt, often referred to as Raman or SRS tilt, where channel powers progressively increase from shorter to longer wavelengths. The Raman tilt depends on the total power of the aggregate signals, and the distribution of that power across the spectrum (i.e., the wavelength locations of the signals) and is modified by the wavelength-dependent fibre loss. These non-linear effects directly impact channel performance metrics, such as optical signal-to-noise ratio (OSNR) and overall quality of transmission (QoT) [5]. Therefore, accurately characterizing the SRS effect is crucial for system engineering and margin design, especially in desegregated systems where end-to-end engineering is not controlled by a single vendor. System vendors or network operators may choose to apply a pre-emphasis tilt on each span in order to ensure uniform OSNR performance across the span. If the wrong tilt is applied or the tilt prediction is erroneous, then the channel performance will vary from design targets. This can require large design margins, resulting in unnecessary costs, and potentially reduced efficiency.

In earlier studies [6], [7], [8], analytical models predicted the effects of SRS with high accuracy for uniformly distributed flat WDM channels. However, splices and other such defects in the fibre can cause variations in a transmission span which, in turn, can lead to notable deviations from initial assumptions. Moreover, WDM channels are unevenly distributed and channel powers fluctuate due to wavelength-dependent power dynamics or modulation-specific engineering rules. These variations can result in substantial inaccuracies in the predictions of analytical models [6], [7], [8]. As such, these models fail to maintain accuracy under dynamic channel loading conditions. Recent advancements in machine learning (ML) have demonstrated excellent performance in predicting Raman tilt across various channel loadings [9]. Despite this progress, ML approaches require extensive data collection to account for variations in fiber types, lengths, and input power levels [10].

A promising alternative is transfer learning (TL) [11], which reduces the need for large datasets by leveraging pre-trained models with similar domain knowledge. In this scenario, the system vendor can train ML models in the lab and then use TL to implement that pre-trained model in a system control framework. This enables the system in operation to collect (reduced) data from the field and apply TL algorithms to get predictions of the Raman tilt which can be used to determine the required pre-emphasis. TL, combined with convolutional neural networks (CNNs), provides a powerful approach for analyzing large datasets. CNNs extract relevant features through convolutional layers while reducing computational complexity through parameter sharing and sparse connections, making them suitable for scalable applications [12]. Most recently [13], [14], CNNs have been used for spectrum prediction in communication networks. In [13], authors design a CNN-based model to accurately predict the spectral changes

978-3-903176-67-6

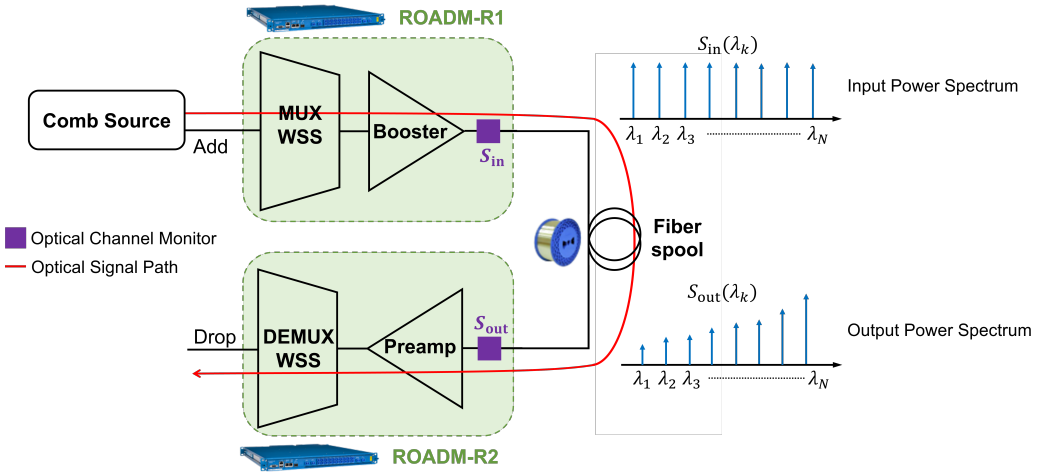

Fig. 1: Topology of the experiments used for data collection in the testbeds (ROADM: reconfigurable optical add-drop multiplexer, WSS: wavelength selective switch).

during the communication process and combat various sources of communication interference. A communication spectrum prediction model is built in [14] to achieve accurate perception and prediction of spectrum resources.

In this study, we apply TL to predict Raman tilt spectra across four distinct fiber configurations in two testbeds. We design a CNN as the base model and transfer knowledge from source configurations to target configurations. Our results demonstrate that the transferred models achieve prediction accuracies that are comparable to those of the base models, with performance primarily constrained by in-field channel monitoring limitations. We evaluate three TL scenarios: (i) between fibers of the same length in different testbeds, (ii) between different input power settings, and (iii) between fibers of varying lengths within the same testbed. We show that the proposed TL-based approach is a promising solution for efficient Raman tilt characterization in optical networks, particularly in scenarios requiring restricted data collection for system adaptation.

II. Experimental Configuration for Data Collection

In this section, we describe the experimental set-up used to collect data for training the proposed CNN-based TL models for the prediction of Raman tilt. We gather data through experiments on four distinct fiber configurations. These experiments use two open-access testbeds: the PAWR COSMOS testbed in Manhattan, New York City [15], and the OpenIreland testbed at Trinity College Dublin, Ireland [16]. The diverse environments in these testbeds allow us to evaluate the model's adaptability to various operational conditions and configurations.

The experimental setup, shown in Fig. 1, includes two reconfigurable optical add-drop multiplexers (ROADMs) connected by spools of standard single-mode fiber. To emulate a wavelength-division multiplexing (WDM) spectrum in the C-band, we use a comb source to generate 90 channels, each spaced at 50 GHz, spanning 191.325 THz–196.075 THz. The comb source sends its output to the add port of the wavelength-selective switch (WSS) in the first ROADM (R1). The WSS selects channels and flattens power levels, ensuring an average channel power of P_0 with deviations limited to ≤ 0.1 dB. Signals travel through the fiber spool to the second ROADM (R2), where the built-in optical channel monitor (OCM) measures the pre-amplifier input. After transmission, the signals pass through a demultiplexing WSS for isolation and analysis. This setup ensures precise control over signal characteristics and facilitates accurate measurements. At the OCMs in R1 and R2, we record the input and output spectra, $S_{in}(\lambda_k)$ and $S_{out}(\lambda_k)$, respectively, under various test conditions which are then used to obtain the Raman tilt as explained below. Here, λ_k denotes the k^{th} wavelength in an N-channel WDM system such that k is an integer in the range $[1, N]$.

The gain evolution across the channel wavelengths is obtained by normalizing the output as $G_k = S_{out}(\lambda_k)/S_{in}(\lambda_k)$. This suppresses the effect of any minor fluctuations in the input spectrum and is, therefore, essential to characterize the stimulated Raman scattering (SRS) effect. We define the Raman tilt as the ratio of the output of the longest wavelength channel to that of the shortest wavelength channel in the WDM spectrum. Specifically, the Raman tilt in an N-channel WDM system is obtained as

$$r_N = \frac{G_N}{G_1}, \tag{1}$$

where G_N and G_1 represent the gains of the longest wavelength (N^{th} channel) and the shortest wavelength (1st channel), respectively. The Raman tilt thus provides a quantitative metric to assess the asymmetry in power distribution induced by SRS across the WDM spectrum.

Table I describes the four fiber configurations. F1 and F3 use L = 25 km fiber spools, while F2 and F4 use L = 50 km spools. The COSMOS testbed (F1 and F2) uses Corning fiber with P_0 = 3.5 dBm, and the OpenIreland testbed (F3 and F4) uses Lucent and Corning fibers with P_0 = 2.7 dBm. Note that we use standard single-mode fiber

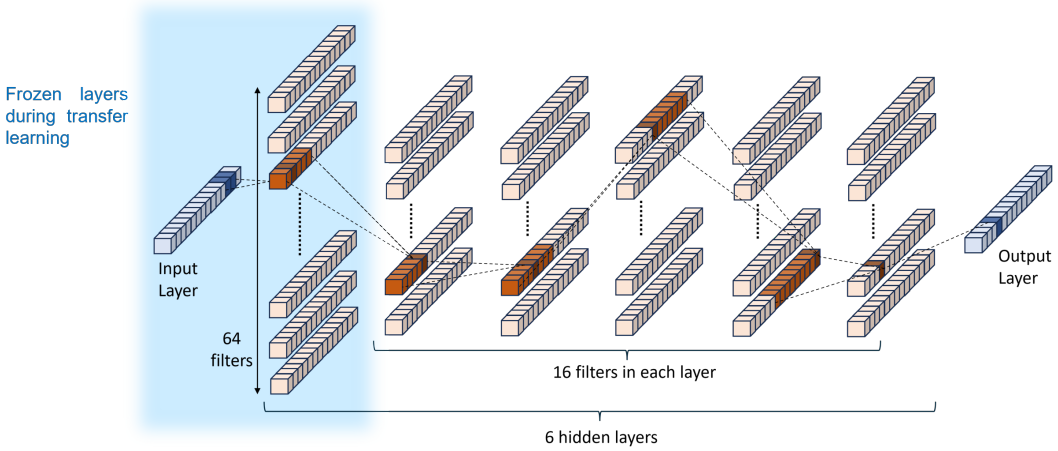

Fig. 2: Proposed CNN model architecture. The small blue and orange cubes represent parameters and neurons, respectively. Kernels are shown in darker shades. The layers frozen during transfer learning are shaded blue.

TABLE I: FIBER CONFIGURATIONS

Configuration	Testbed	Fiber Length L (km)	Input Power P_0 (dBm)
F1	COSMOS	25	3.5
F2	COSMOS	50	3.5
F3	OpenIreland	25	2.7
F4	OpenIreland	50	2.7

TABLE II: CNN ARCHITECTURE

Parameter	Value
Input features	Input power spectrum
Output features	Output power spectrum
Optimization algorithm	Adam optimizer
Activation function	ReLu
No. of hidden layers	6
Kernel sizes	3, 5, 7, 9, 11
Dilation rates	1, 2, 4, 8, 16
Loss function	Mean squared error (MSE)
Learning rate	0.001 with exponential decay

(SSMF) in all four configurations. For each configuration, we systematically vary the number of active WDM channels, $N \in \{2, 5, 10, 20, 30, \ldots, 80, 90\}$, loaded at the input of the fiber. Here, the channels are randomly distributed across the complete spectral band. For $N < 90$, the channels can occupy any N out of 90 possible spectral locations. Hence, for each value of N, there are ${}^{90}C_N$ different possible channel distributions, where ${}^{x}C_y$ is the binomial coefficient defined as ${}^{x}C_y = x!/y!\,(x-y)!$. Moreover, in general, the channel spacing is uneven although the minimum channel spacing is bounded at 50 GHz. Furthermore, the total spectral bandwidth is fixed. Using the experimental set-up in Fig. 1, we generate 100 unique channel distributions for each value of N, and record the observations for all four fiber configurations, generating 4,400 spectra in machine-readable JSON format. Note that for $N = 90$, all experiments have the same spectral distribution of channels.

III. CNN-BASED TRANSFER LEARNING (TL) MODEL

In this section, we elucidate the proposed model for Raman tilt prediction in WDM systems. We design a CNN-based source model to predict the Raman gain profile for each fiber configuration. The model input consists of the normalized power levels of individual wavelength channels at the fiber input. The power spectrum, measured in the decibel scale, is converted to the linear scale and normalized using min-max scaling. This input layer feeds into six hidden convolutional layers, each designed to effectively capture the inherent patterns in the data. The hidden layers are configured with filters in the sequence 64/16/16/16/16, ensuring that the model maintains computational efficiency while retaining the ability to extract meaningful features.

The hidden layers employ progressively increasing kernel sizes of 3, 5, 7, 9, 11, and 13, combined with exponentially growing dilation rates of 1, 2, 4, 8, 16, and 32, respectively. These settings expand the receptive field of the network, allowing it to capture long-range dependencies within the input data. To handle out-of-bound elements, we apply consistent zero-padding across the convolutional layers. Each layer is followed by layer normalization to stabilize training and improve convergence, while the rectified linear unit (ReLU) activation function introduces non-linearity and ensures efficient gradient flow. The CNN source model is trained for 250 epochs. This choice is informed by observing that the loss curve reaches its lowest point and stabilizes at this stage with minimal fluctuation. The loss curve becomes unstable if the training is extended beyond 250 epochs, making this a robust stopping point for reliable performance. To further optimize the training process, we adopt an exponential learning rate decay schedule. The initial learning rate is set to 0.001, which decays by a factor of 0.95 every 1000 training steps. By applying this decay, the learning rate is adjusted smoothly, allowing the model to take progressively smaller steps as it approaches convergence. This ensures that the optimization becomes increasingly conservative in deeper stages of training, reducing the likelihood of skipping the global minimum. The mean squared error (MSE) is used as the loss function to optimize predictions. The details of the proposed CNN architecture are summarized in Table II. For each value of N, representing

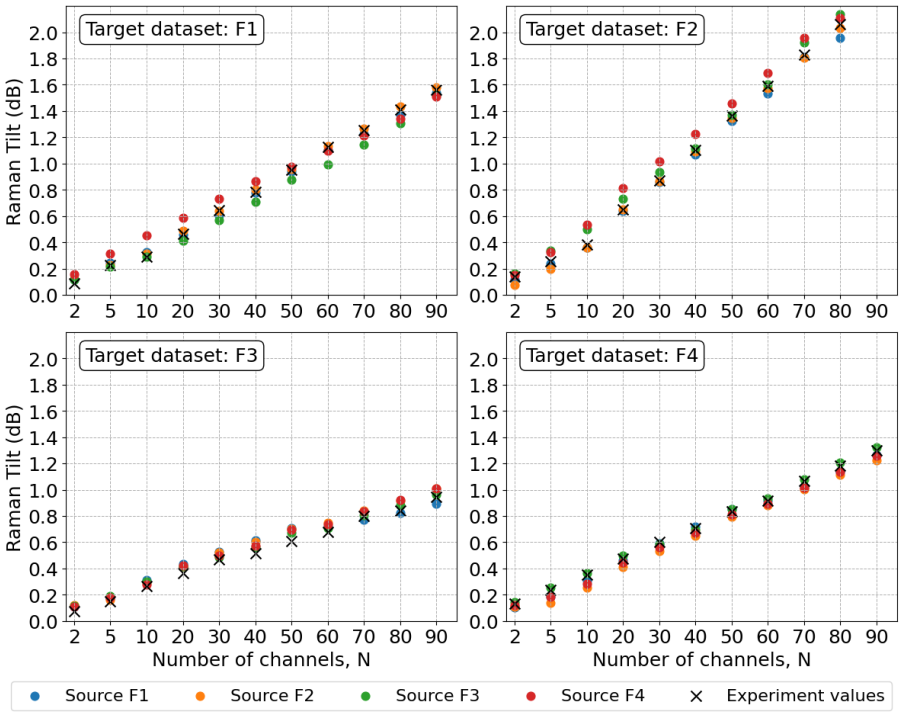

Fig. 3: Raman tilt obtained from experimental observations and predictions for different target fiber configurations using TL and base model of the source fiber configurations. The fiber configurations are described in Table I.

the number of active WDM channels, we split the observed dataset into training and testing subsets with an 80:20 ratio.

To minimize data collection efforts for new configurations, we employ TL to adapt the source model to target datasets. The input layer and the first hidden layer, comprising a total of 5,392 parameters, are frozen to retain the pre-trained weights from the base model, preserving the foundational feature extraction capabilities. The remaining layers are re-initialized using Kaiming initialization to fine-tune the model for the specific characteristics of the target dataset. During the transfer process, only a portion of the target dataset is utilized for training, allowing the model to effectively adapt to the specific features of the target configuration while being retrained using the same training parameters for an additional 100 epochs.

IV. Results

In this section, we discuss the results of our proposed prediction model. We present the Raman tilt values obtained using the experimentally recorded spectra as explained in Section II and compare them with the corresponding predictions made by the proposed CNN-based model.

We employ TL to predict the output spectra of a *target* fiber configuration using the CNN-based model of the *source* fiber configuration. We obtain the Raman tilt using the output spectra obtained from experiments and from CNN/TL predictions, i.e., $r_{i,N}$ and $\hat{r}_{i,N}$, respectively. In Fig. 3, we depict the Raman tilt predictions for different source and target fiber configurations. The fiber configurations are described in Table I. We observe that, in all configurations, the Raman tilt predictions using the proposed CNN are very close to the corresponding observed values. Moreover, for a given value of N, the Raman tilts obtained for distinct target fiber configurations vary due to the different fiber lengths and input power levels in these configurations. For example, at $N = 40$, the predicted Raman tilt values for target configurations F1, F2, F3 and F4 are centred around 0.8 dB, 1.1 dB, 0.6 dB and 0.7 dB, respectively.

We now evaluate the performance of the prediction model by computing the mean absolute error (MAE) in the predicted values when compared with the actual observed values in the experiments. For a given N, the MAE is calculated as

$$\mathcal{E}_N = \frac{1}{K}\sum_{i=1}^{K} |r_{i,N} - \hat{r}_{i,N}| \tag{2}$$

where K is the number of experiments conducted for each N, and $r_{i,N}$ and $\hat{r}_{i,N}$ are the observed and predicted values of Raman tilt, respectively, obtained in the i^{th} experiment for N number of active channels, using Eqn. 1. As discussed in Sections II and III, for this study, we fix K = 20 (for the test dataset) and $N \in \{2, 5, 10, 20, 30, \ldots, 80, 90\}$. We use MAE matrices to depict the error performance under different prediction scenarios. In each MAE matrix, the non-diagonal elements $(i, j), i \neq j$, correspond to the transferred model with Fi and Fj being the source and target fiber configurations, respectively, where $i, j \in \{1, 2, 3, 4\}$. The diagonal elements (i, i) correspond to the component-level CNN-based model without TL (i.e., when the same configuration is being used as the source and the target).

Fig. 4 illustrates the MAE matrices across varying spectrum occupancy levels: sparsely-loaded (N = 5), moderately-loaded

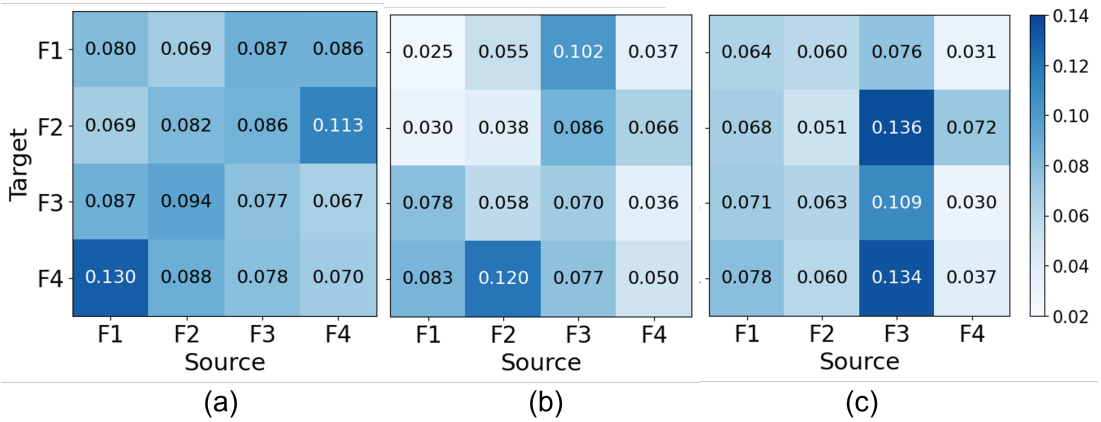

Fig. 4: Mean absolute error (MAE) matrices for Raman tilt prediction when the number of active WDM channels is (a) $N = 5$, (b) $N = 40$, and (c) $N = 90$. In each MAE matrix, the non-diagonal elements $(i, j), i \neq j$, correspond to the transferred model with Fi and Fj being the source and target fiber configurations, respectively, where $i, j \in \{1, 2, 3, 4\}$. The diagonal elements (i, i) correspond to the component-level CNN-based model without TL. The fiber configurations are described in Table I.

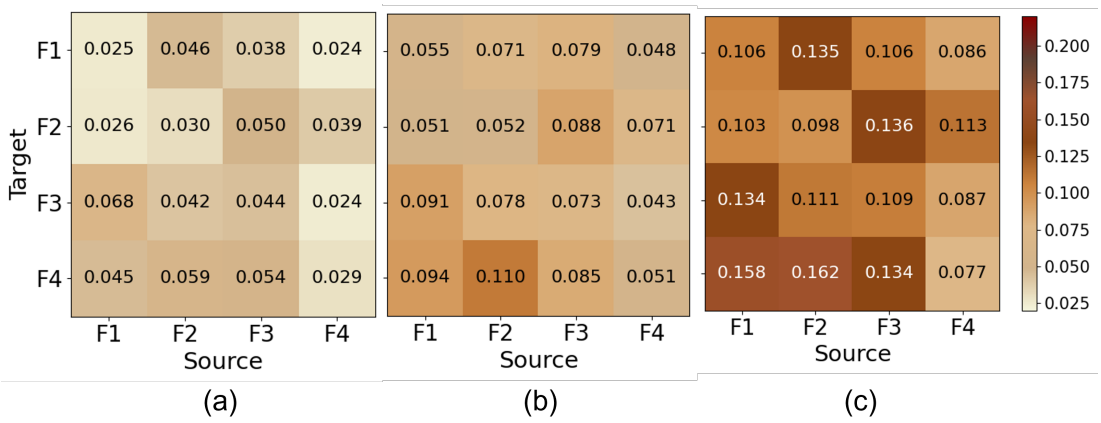

Fig. 5: Mean absolute error (MAE) matrices for Raman tilt prediction depicting the (a) minimum, (b) average, and (c) maximum values across all values of N. In each MAE matrix, the non-diagonal elements $(i, j), i \neq j$, correspond to the transferred model with Fi and Fj being the source and target fiber configurations, respectively, where $i, j \in \{1, 2, 3, 4\}$. The diagonal elements (i, i) correspond to the component-level CNN-based model without TL. The fiber configurations are described in Table I.

(N = 40), and fully-loaded (N = 90). For $N = 5$ and $N = 40$, the highest MAEs are 0.13 dB and 0.12 dB, respectively, obtained when F4 is the target configuration with F1 and F2 as the source configurations, respectively. For $N = 90$, the highest MAE (0.136 dB) is incurred in the predictions for the F2 configuration when transferring from the F3 configuration. Note that these errors are obtained when 70% of the data from the target dataset is utilized for training during the transfer process. Using smaller portions of the target dataset for TL aggravates the peak errors in prediction. Further investigations are needed to determine methods that would allow for greater reduction in the field-collected data.

In Fig. 5, we plot the minimum, average, and maximum MAEs across all values of N with respect to the four fiber configurations (described in Table I). Here, for each source-target combination (represented by a cell in the MAE matrix), we can observe the range of MAEs and the average error performance across all values of N. For example, when F1 and F4 are the source and target configurations, respectively, the MAE across all values of N ranges from 0.045 dB to 0.158 dB with an average at 0.094 dB. When averaged across all N (Fig. 5b), the best error performance by the proposed TL model is obtained for the F3 configuration when transferred from the F4 configuration with an MAE of 0.043 dB. Overall, the prediction MAE between target and source configurations ranges from 0.024 dB to 0.162 dB. We also show the effect of variation in N on the prediction performance. Fig. 6 illustrates the distribution of MAE in prediction for different target fiber configurations, showing both the mean and standard deviation. We observe that the MAE does not vary much as the number of channels increases. Better performance from base models is consistently witnessed, and the MAE is largely reduced to below 0.1 dB, in many cases approaching the $\pm$ 0.05 dB resolution of the OCM (least count 0.1 dB) available in the ROADMs for in-service data collection, as depicted in Fig. 1.

V. Conclusion

In this study, we design and evaluate a CNN-based transfer learning (TL) model to predict Raman tilt spectra across four different fiber configurations in two geographically distinct testbeds. The proposed CNN model demonstrates high prediction accuracy, achieving a mean absolute error (MAE)

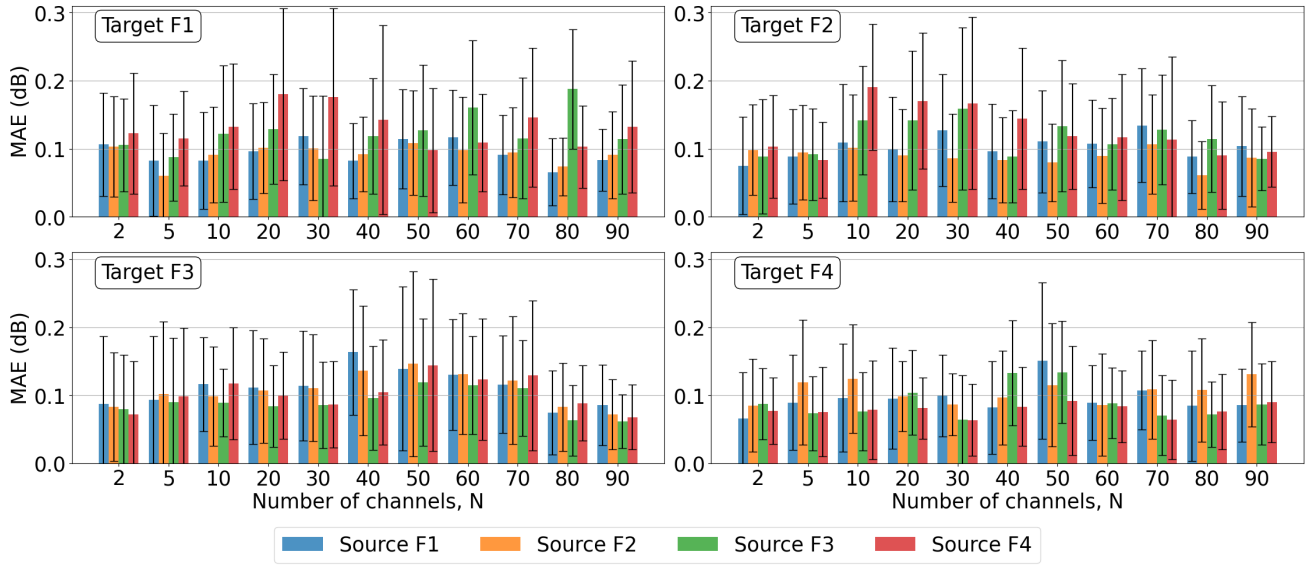

Fig. 6: Mean absolute error (MAE) in predictions for different target fiber configurations using TL and base model of the source fiber configurations. The fiber configurations are described in Table I.

as low as 0.024 dB and no higher than 0.162 dB, with errors primarily limited by measurement inaccuracies in the experimental setup. Our findings highlight the effectiveness of TL for predicting Raman tilt under varying fiber configurations, input power levels, and fiber lengths. Notably, the prediction performance is relatively consistent across different spectrum occupancy levels, with MAE values remaining largely unaffected by the number of active channels. This robustness underscores the adaptability of the CNN-based TL approach to various operational conditions. Additionally, the results reveal that the CNN-based base models consistently outperform transferred models. However, transferred models still achieve acceptable prediction accuracy, while reducing the data collection requirements for new target configurations by 30%. This work serves as a first step towards exploring the effectiveness of TL in prediction of Raman tilt with acceptable accuracies. Nevertheless, further research in this direction is needed to discover ways for achieving greater reduction in data collection for new target configurations while maintaining acceptable prediction errors. These efforts aim to reduce reliance on extensive in-field data measurements, thereby facilitating scalable and cost-effective optical network management.

Acknowledgments

This work was supported in part by Science Foundation Ireland under grant numbers 13/RC/2077 P2 and 22/FFP-A/10598; National Science Foundation under grant numbers CNS-1827923, OAC-2029295, CNS-2112562, and CNS-2330333; and the European Union under MSCA grant number 101155602.

References

[1] A. Lord, "The future of optical transport: Architectures and technologies from an operator perspective," in *Proc. Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–18, San Diego, USA, 2022.

[2] Y. Pointurier, "Design of low-margin optical networks," *IEEE/Optica Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. A9–A17, 2017.

[3] A. Mahajan, K. Christodoulopoulos, R. Martinez, S. Spadaro, and R. Munoz, "Machine learning assisted EDFA gain ripple modelling for accurate QoT estimation," in *Proc. European Conference on Optical Communication (ECOC)*, pp. 1–4, Dublin, Ireland, 2019.

[4] D. Christodoulides and R. Jander, "Evolution of stimulated Raman crosstalk in wavelength division multiplexed systems," *IEEE Photonics Technology Letters*, vol. 8, no. 12, pp. 1722–1724, 1996.

[5] R. D'Ingillo *et al.*, "Deep learning gain and tilt adaptive digital twin modeling of optical line systems for accurate OSNR predictions," in *Proc. International Conference on Optical Network Design and Modeling (ONDM)*, pp. 1–3, Madrid, Spain, 2024.

[6] M. Zirngibl, "Analytical model of Raman gain effects in massive wavelength division multiplexed transmission systems," *Electronics Letters*, vol. 34, no. 8, pp. 789–790, 1998.

[7] S. Bigo, S. Gauchard, A. Bertaina, and J.-P. Hamaide, "Experimental investigation of stimulated Raman scattering limitation on WDM transmission over various types of fiber infrastructures," *IEEE Photonics Technology Letters*, vol. 11, pp. 671–673, 1999.

[8] F. Vanholsbeeck, S. Coen, P. Emplit, M. Haelterman, and T. Sylvestre, "Raman-induced power tilt in arbitrarily large wavelength-division-multiplexed systems," *IEEE Photonics Technology Letters*, vol. 17, pp. 88–90, 2005.

[9] R. Raj, S. Xie, Z. Wang, T. Chen, and D. Kilper, "Machine learning-based Raman tilt prediction in a ROADM transmission system," in *Proc. European Conference on Optical Communications (ECOC)*, pp. 1504–1507, Glasgow, Scotland, 2023.

[10] G. Liu *et al.*, "Hierarchical learning for cognitive end-to-end service provisioning in multi-domain autonomous optical networks," *IEEE/Optica Journal of Lightwave Technology*, vol. 37, no. 1, pp. 218–225, 2019.

[11] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[12] S. Muthu, R. Tennakoon, R. Hoseinnezhad, and A. Bab-Hadiashar, "A survey of CNN-based techniques for scene flow estimation," *IEEE Access*, vol. 11, pp. 99289–99303, 2023.

[13] Z. Tian, D. Yu, Y. Bai, S. Lei, and Y. Wang, "Application of spectrum state prediction method based on CNN-LSTM network in communication interference," *IEEE Access*, vol. 11, pp. 93538–93550, 2023.

[14] L. Yuan, L. Nie, and Y. Hao, "Communication spectrum prediction method based on convolutional gated recurrent unit network," *Nature Scientific Reports*, vol. 14, no. 1, p. 8959, 2024.

[15] T. Chen *et al.*, "A software-defined programmable testbed for beyond 5G optical-wireless experimentation at city-scale," *IEEE Network*, vol. 36, no. 2, pp. 90–99, 2022.

[16] CONNECT, "OpenIreland testbed, funded by Science Foundation Ireland," 2024.