POLITECNICO DI TORINO
Repository ISTITUZIONALE

Spectral and Spatial Power Evolution Design With Machine Learning-Enabled Raman Amplification

(Article begins on next page)

24 November 2024

Spectral and spatial power evolution design with machine learning-enabled Raman amplification

Mehran Soltani (iD), Francesco Da Ros (iD), Andrea Carena (iD), Darko Zibar (iD)

Abstract—We present a machine learning (ML) framework for designing desired signal power profiles over the spectral and spatial domains in the fiber span. The proposed framework adjusts the Raman pump power values to obtain the desired two-dimensional (2D) profiles using a convolutional neural network (CNN) followed by the differential evolution (DE) technique. The CNN learns the mapping between the 2D profiles and their corresponding pump power values using a data-set generated by exciting the amplification setup. Nonetheless, its performance is not accurate for designing 2D profiles of practical interest, such as a 2D flat or a 2D symmetric (with respect to the middle point in distance). To adjust the pump power values more accurately, the DE fine-tunes the power values initialized by the CNN to design the proposed 2D profile with a lower cost value. In the fine-tuning process, the DE employs the direct amplification model which consists of 8 bidirectional propagating pumps, including 2 second-order and 6 first order, in an 80 km fiber span. We evaluate the framework to design broadband 2D flat and symmetric power profiles, as two goals for wavelength division multiplexing (WDM) system performing over the whole C-band. Results indicate the framework's ability to achieve maximum power excursion of 2.81 dB for a 2D flat, and maximum asymmetry of 14% for a 2D symmetric profile.

Index Terms—Inverse system design, machine learning and optimization, power evolution design, Raman amplification.

I. INTRODUCTION

Distributed Raman amplification provides the opportunity to design the signal power evolution jointly in spectral and spatial domains. This amplification scheme offers variety of attractive features such as low noise properties [1], [2], broad range of gain due to multi-pumping configuration and flexibility in gain and power evolution design [3], [4]. One of the main challenges with Raman amplifier design is to select the pump powers and their frequency values to achieve desired signal power profiles in the fiber span. Recently, most of the focus has been on designing a desired gain spectra at the receiver side using machine learning and optimization techniques [4]–[7]. In [5] a neural network (NN) is combined with genetic algorithm to design flat gain profiles. The authors of [4], [6] propose a NN model to learn the inverse mapping between the gain profiles and the pump powers and wavelengths values. Additionally, an extra NN model is proposed in [4], [6] to fine-tune the pump parameter values through back-propagation. In [7], a differentiable Raman amplification model is presented in the training procedure of NN to predict the pump parameters to target a group of flat or tilted gains.

M. Soltani, F. Da Ros and D. Zibar are with the Department of Photonics engineering, Technical university of Denmark (DTU Fotonik), Denmark
A. Carena is with the Dipartimento di Elettronica e Telecomunicazioni at Politecnico di Torino, Italy

Besides gain spectra shaping, designing a 2D signal power evolution profile jointly in spectrum and space is a beneficial way to satisfy some of the goals in an optical communication system like enhancing the signal-to-noise ratio (SNR) and mitigation of nonlinear impairments [8]. For example, distribution of the power evenly in frequency and distance along the fiber resembles a link with effective zero attenuation and results in the minimum accumulated amplified spontaneous emission (ASE) noise [9]–[11]. Additionally, a zero-loss medium is a theoretical prerequisite for the transmission based on Nonlinear Fourier Transform (NFT) which relies on analytically solving a lossless nonlinear Schrödinger equation (NLSE) [12], [13]. Alternatively, a possibly less challenging and more realistic target profile in distance for nonlinearity compensation is a symmetric profile. A symmetric power profile with respect to the middle point in distance is required to maximize the performance of an optical phase conjugation (OPC) [14] in combating the nonlinear impairments in the system [15], [16].

It has been numerically and experimentally proven that a desired 2D signal power evolution can be achieved using high-order Raman pumps in a bidirectional propagating scheme [17]. Most common Raman amplification setups for designing the power evolution in fiber distance are made by combining second-order and first-order pumps in a bidirectional propagating scheme [17], or by using Raman pumps combined with fiber bragg gratings (FBGs) as reflectors [9]. In [9], [11], an amplification structure based on second-order Raman pumps with FBGs has been presented to provide a quasi-lossless transmission medium by implementing an ultra long Raman fiber laser. Moreover, in [18], numerical optimization of narrow-band signal power asymmetry for OPC has been addressed by considering different amplification setups using first-order and second-order pumps.

Most of the works presented for designing signal power profiles have addressed single-channel narrow-band signals, with a few interesting works approaching a multi-channel quasi-lossless signal transmission [11], [19]–[21]. In these papers, the authors present amplification setups including second-order pumps combined with the FBG reflectors to turn the span into an ultra-long laser, aiming to achieve a lossless transmission in frequency and distance. These systems relied on either optimizing the pumps heuristically [11], [19], or optimizing forward pumps and then adjusting the backward pumps to match the desired gain for the central channel [20], or assuming equal power for backward and forward pumping, thus simplifying the optimization problem [21]. These approaches become challenging when the aim is to
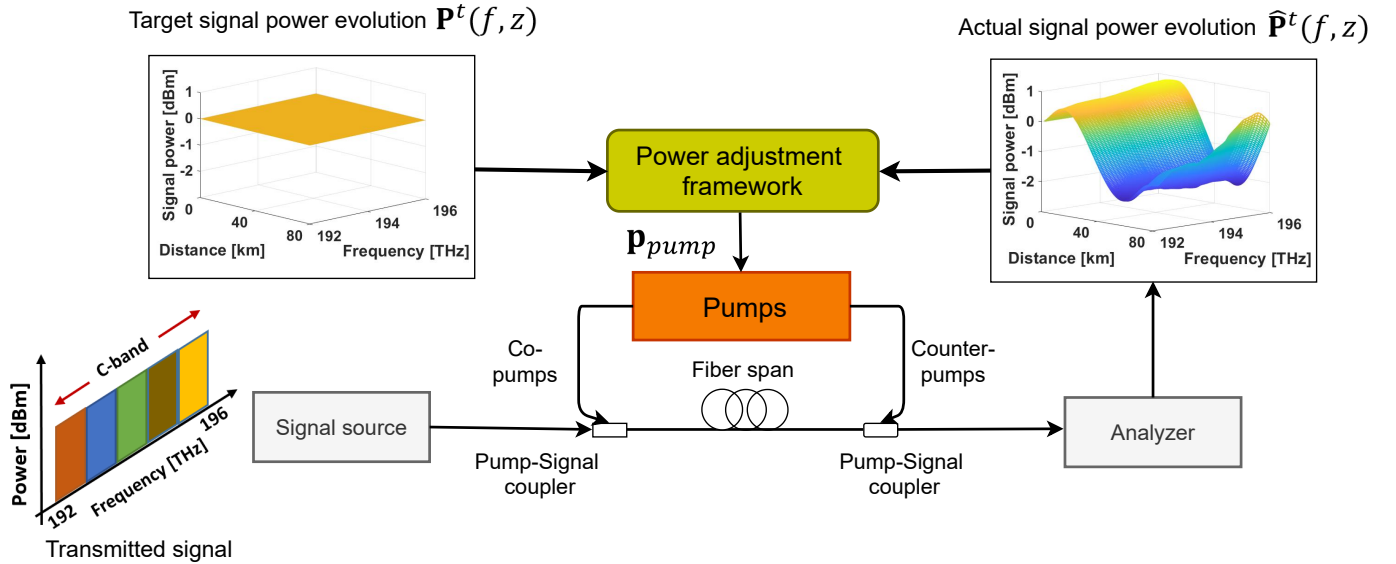
Fig. 1. Inverse system model applied to adjust the Raman pump powers values in an amplification to design a desired 2D signal power profile.

provide a general framework to design a desired 2D power profile of practical interest.

In this paper we present a ML-based framework to design a desired 2D signal power evolution in wide-band spectrum and along the fiber span by adjusting Raman pump powers values. The proposed framework extends our presented CNN model [22] which approximates the mapping between the 2D power profiles and their corresponding Raman pump powers values. The CNN model is trained on a data-set generated by exciting the amplification model with randomly selected pump powers values. In spite of this, its performance is not accurate for designing specific 2D target profiles. In addition, 2D profiles such as a 2D flat or a 2D symmetric have specific and non-differentiable cost functions, which cannot be easily approached with learning-based models like the CNN. To obtain more accurate pump powers values for these profiles, we consider the predicted values by the CNN as an initial solution and apply differential evolution (DE) as a gradient-free fine-tuning technique. The proposed CNN-assisted DE framework uses a direct functional description of the amplification model and fine-tunes the pump power values to design a 2D profile with lower cost value based on the desired objective. The framework is evaluated on designing 2D flat and symmetric power profiles, as two highly addressed targets of practical interest in the literature [9], [18].

Alternative to the first-order Raman amplification setup in [22], we propose a higher-order pumping scheme to enable the framework to approach desired 2D profiles. More specifically, we use a bidirectional amplification setup with two second-order and six first order pumps (one second-order and three first-order for each propagation direction) to cover the whole C-band without adding system complexity such as FBGs [9]. Moreover, we assume fixed wavelengths values for the Raman pumps same as in the experiments, where the pump module usually allows to only set the power values. Therefore, the

2D profile design problem will be confined to the prediction of Raman pump powers values. Nevertheless, unlike the presented approaches in [20], [21], the proposed framework has the flexibility to fully optimize the pumps independently, and to be applied to a different amplification setup without making any assumptions regarding its physical properties.

The remainder of the paper is organized as follows: In section II, we describe the proposed amplification structure followed by the CNN-assisted DE framework for 2D power evolution design problem. In section III, first, the performance of the inverse CNN model is evaluated on designing achievable power profiles. Afterward, the proposed CNN-assisted DE is applied to design 2D flat and symmetric power profiles. Finally, we conclude the paper in section IV.

II. 2D SIGNAL POWER EVOLUTION DESIGN

In this section, first, we will describe a general view of a Raman amplification scheme which utilizes a Raman pump power adjustment framework to design a 2D power evolution profile. Afterward, we will present our suggested framework to design 2D spectral-spatial profiles in more details.

A. Raman pump power adjustment framework

Fig. 1 illustrates a schematic of a fiber link in which a target 2D signal power profile in spectrum and fiber distance is designed by adjusting the Raman pump power values. In this scheme, a 2D target signal power evolution $\mathbf{P}^t(f, z)$ is used as the input to the framework which predicts the Raman pump power values $\mathbf{p}_{pump}$, including both co- and counter-propagating pumps. The predicted pump powers values are then applied by the Raman pumps into the fiber span where a WDM frequency comb signal with flat spectra in C-band is propagated.

Due to the stimulated Raman scattering (SRS) phenomenon, the fiber is utilized as the gain medium. Consequently, the

signal power evolution in both frequency and distance is tailored by the SRS and a two dimensional matrix referred to as the actual 2D signal power evolution $\hat{\mathbf{P}}^t(f, z)$ is calculated. Afterward, $\hat{\mathbf{P}}^t(f, z)$ is compared to $\mathbf{P}^t(f, z)$ to calculate the corresponding cost value based on the desired objective. The pump power adjustment framework evaluates the cost value and applies an updated set of Raman pump power values to the system aiming to produce a new 2D signal power profile with a lower cost value. This process, which is referred to as "fine-tuning", continues until a certain convergence criteria such as maximum number of iterations or a certain threshold for cost value is satisfied.

As detailed in section I, the fine-tuning models presented in the literature for Raman amplification design have addressed gain spectrum shaping at the receiver [4], [6]. These frameworks employ a NN model to approximate the "direct models" of the amplification system. The direct model is used to evaluate the gain spectrum given the pump parameters values. The cost function in these frameworks is defined as mean square error (MSE) which is differentiable with respect to the model parameters and enables to fine-tune the pump parameters values through back-propagation.

Addressing the target 2D power evolution profiles of practical interest, cost functions for each target may be non-differentiable, and also target-dependent. For instance, for a 2D flat or a 2D symmetric profile, the objective is to minimize the power excursion (also called flatness) or the asymmetry in distance , respectively [11], [18], [23]. For these 2D targets, it is difficult or even may not be possible to calculate the gradient of the cost function with respect to the model parameters. Additionally, it is shown that mean square error (MSE), despite its simplicity and mathematically convenient form, is not adequate for the assessment of data resembled as 2D images [24]. Considering these issues, we model the pump power adjustment process for designing target 2D practical power profiles as a gradient-free global optimization problem. The proposed framework has the flexibility to design 2D practical power profiles with different corresponding forms of objectives.

B. CNN-assisted DE for 2D power evolution design

In the proposed framework, we choose DE algorithm, known as an evolutionary-based technique to adjust the pump powers values. DE is a gradient-free technique with the robustness and the flexibility to capture solutions of complex optimization problems [25]. More importantly, it has been successfully applied to a wide variety of problems due to its strong global search capability, easy implementation and quick convergence [26].

Generally, evolutionary algorithms like DE start with a population of individuals (as the initial solutions) and try to improve each particle of the population in an iterative process to achieve a point with minimum cost value. In the absence of a prior information about the initial solution, they always start with a random guess in the parameter space [27]. Considering this, a requirement of the DE for the proposed 2D power evolution design is to define the search region of pump power

values by setting their lower-bound $\mathbf{p}_{LB}$ and the upper-bound $\mathbf{p}_{UB}$ values. A possible way of specifying $\mathbf{p}_{LB}$ and $\mathbf{p}_{UB}$ is to set them as the minimum and maximum possible values of pump powers, respectively. This approach is proposed generally when there is no prior information provided about the possible sub-region of the pump power values including the point with minimum cost value. With this blind search approach, the model would be liable to slow convergence, and moreover, plunging into local minimum. This issues are more dominant when the number of pumps, being the dimension of the space to search, increases [28].

We improve our chance to approach a more reliable and accurate solution for the proposed 2D profile design problem using a ML-based model. The role of the ML model is to provide an initial information regarding a sub-region in the space of pump power values to initialize the DE population. For this purpose, we use our previously presented CNN model in [22] to approximate the mapping between the 2D power profiles and their corresponding pump power values.

The CNN model in [22] is a neural network trained based on a data-set generated by random selection of the pump powers values in the proposed amplification setup. It consists of two networks trained end-to-end, a feature extraction and a regression network. The feature extraction network consists of three CNN layers each followed by pooling layers, while the regression network is made up of 2 fully-connected (FC) layers. The role of the first network is to extract the informative and low-dimensional features of the 2D profiles. In the meantime, the second network maps the extracted features to the corresponding pump parameter values. The training process of the CNN aims to update the network weights to minimize the MSE between the true pump powers values and the predicted ones. The CNN has shown a good performance in designing achievable 2D power evolution profiles which are recorded by random selection of the pump powers values in the amplification setup.

Compared to the achievable set of 2D profiles, the profiles of practical interest are more challenging to obtain due to physical limitations of the amplification systems. Additionally, some practical profiles have strict objectives such as minimum power excursion or minimum asymmetry which cannot be easily approached with the CNN model trained to minimize a loss function such as the MSE. Therefore, in the proposed framework, we employ the CNN to provide a set of pump powers values to be used as an initial solution. This solution gives a prior information about a space region of pump power values where a better solution to the problem falls inside. By initializing the DE population based on the CNN results, we improve the convergence speed and also the quality of the final solution. Considering this, the fine-tuning process is performed with parameter constraints on a relatively tight space region around the set of values predicted by the CNN model.

Alternative to the approximate NN-based direct models used for fine-tuning in [4], [6], we employ the DE combined with the numerically provided direct model of the Raman amplification scheme [1]. We refer to the functional model described between the pump powers values and their corresponding 2D power profiles as Raman solver $\mathcal{R}(\cdot)$. The 2D power evolution
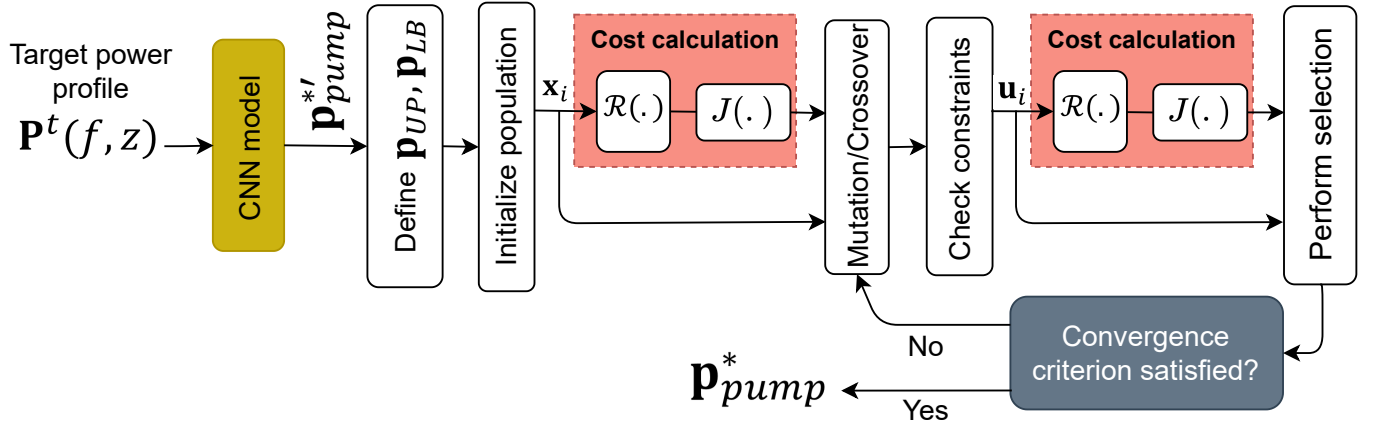
Fig. 2. Block diagram of the proposed CNN-assisted DE method used for pump power prediction based on a specific 2D target power profile.

generated by the Raman solver with a given set of pump power values $\mathbf{p}_{pump}$ is defined as:

$$\mathbf{P}(f, z | \mathbf{p}_{pump}) = \mathcal{R}(\mathbf{p}_{pump}) \qquad (1)$$

Indeed, the $\mathcal{R}(\cdot)$ function is a set of nonlinear differential equations describing the Raman effect on the signal power evolution [1]. This set of equations is generally considered as a boundary value problem (BVP) which is solved numerically using ordinary differential equation (ODE) techniques [29].

Diagram of the proposed Raman pump power adjustment scheme referred to as "CNN-assisted DE", is shown in Fig. 2. First, the 2D target power profile $\mathbf{P}^t(f, z)$ is used as the input to the trained CNN model to predict the corresponding Raman pump power values as the initial solution $\mathbf{p}_{pump}^{*'} = [p_1, ..., p_n]$, where n is the number of pumps. Afterward, the constraints $\mathbf{p}_{LB} = \mathbf{p}_{pump}^{*'} - \mathbf{\Delta}_p \cdot \mathbf{p}_{pump}^{*'}$ and $\mathbf{p}_{UB} = \mathbf{p}_{pump}^{*'} + \mathbf{\Delta}_p \cdot \mathbf{p}_{pump}^{*'}$ are defined, where $\mathbf{\Delta}_p = [\Delta_1, \Delta_2, ..., \Delta_n]$ is a hyper-parameter vector which consists of n number of unit-less scalar values. Each element of $\mathbf{\Delta}_p$ is multiplied by its corresponding initially pump power value in $\mathbf{p}_{pump}^{*'}$. The resulting number for each pump power is considered as its amount of deviation and the constraints $\mathbf{p}_{LB}$ and $\mathbf{p}_{UB}$, accordingly.

Once the constraints are defined, an initial population with N_p number of individuals is generated from the uniform distribution over the specified ranges. Each individual in the population is represented by a vector $\mathbf{x}_i = (x_i^1, x_i^2, ..., x_i^n), i = 1, ..., N_p$ as a set of pump power values, which is a candidate for the fine-tuning solution. All of the individuals are passed through the Raman solver in Eq. 1 and their corresponding 2D profiles are generated. The resulting 2D profile for each candidate is used further for the calculation of the corresponding pre-defined cost value. The cost value of the ith individual $J(\mathbf{x}_i)$ can be defined either as an error between the 2D target $\mathbf{P}^t(f, z)$ and the resulting 2D profile $\hat{\mathbf{P}}(f, z | \mathbf{x}_i) = \mathcal{R}(\mathbf{x}_i)$, or based on a metric measured for the resulting 2D profile such as its power excursion or asymmetry value.

For each generated individual $\mathbf{x}_i$, an iterative DE process including mutation, crossover and evaluation is performed. First, three individuals $\mathbf{x}_{r_1}$, $\mathbf{x}_{r_2}$ and $\mathbf{x}_{r_3}$ are selected randomly

from the population where $r_1 \neq r_2 \neq r_3 \neq i$, and a mutation is applied to generate a donor vector $\mathbf{v}_i$ with the same size as $\mathbf{x}_i$. The donor vector $\mathbf{v}_i$ aims to increase the diversity of the population in order to avoid plunging into a local minimum, and it is defined as the following:

$$\mathbf{v}_i = \mathbf{x}_{r_1} + F \cdot (\mathbf{x}_{r_2} - \mathbf{x}_{r_3}) \qquad (2)$$

where $F \in [0, 1]$ is the mutation factor, which controls the diversity of the population.

After the mutation, the crossover process develops trial vector $\mathbf{u}_i$ by combining the elements of the donor vector $\mathbf{v}_i$ and the current target vector $\mathbf{x}_i$ as follows:

$$\mathbf{u}_i^j = \begin{cases} v_i^j, & \text{if } r_i^j \leq CR \text{ or } j = j_{rand} \\ x_i^j, & \text{otherwise} \end{cases} \qquad (3)$$

where u_i^j, v_i^j and x_i^j represent the jth element of $\mathbf{u}_i$, $\mathbf{v}_i$ and $\mathbf{x}_i$, respectively. CR is the crossover probability, $r_i^j \sim U(0, 1)$ is a uniform distribution which is generated for each j, and $j_{rand} \in 1, 2, ..., N_p$ is a random integer used to ensure that $\mathbf{x}_i \neq \mathbf{u}_i$. After performing the mutation and the crossover, the trial vector is checked if it follows the constraints $\mathbf{p}_{LB}$ and $\mathbf{p}_{UB}$. Afterward, $\mathbf{u}_i$ is passed through the Raman solver function to generate the corresponding power evolution profile $\mathbf{P}(f, z | \mathbf{u}_i) = \mathcal{R}(\mathbf{u}_i)$ and calculate the cost value J, accordingly. The cost values of the trial vector $J(\mathbf{u}_i)$ and the target vector $J(\mathbf{x}_i)$ will be compared and the one with lower cost value will remain in the population. The selection process can be formulated to form the new particle $\mathbf{x}_{i_{new}}$ as follows:

$$\mathbf{x}_{i_{new}} = \begin{cases} \mathbf{u}_i, & \text{if } J(\mathbf{u}_i) < J(\mathbf{x}_i) \\ \mathbf{x}_i, & \text{otherwise} \end{cases} \qquad (4)$$

The aforementioned process continues further until a convergence criteria is satisfied either by reaching the maximum number of iterations, or by achieving a cost function threshold. In this paper we consider the maximum possible number of cost function evaluations $MaxEv$ as the convergence criteria. The implementation procedure of the proposed CNN-assisted DE optimization scheme is summarized in Algorithm 1.

Algorithm 1 CNN-assisted DE

Input: Target power profile $\mathbf{P}^t(f,z)$, $\Delta_p = [\Delta_1, \Delta_2, ..., \Delta_{N_p}]$, DE parameters: {$N_p$: Population size, CR: Crossover probability, F: Mutation factor, $MaxEv$: Maximum number of cost function evaluations}

Output: Pump power values $\mathbf{p}^*_{pump}$

Optimization procedure:

1: $\mathbf{p}^{*\prime}_{pump} = CNN(\mathbf{P}^t(f,z))$

2: set $\mathbf{p}_{LB} = \mathbf{p}^{*\prime}_{pump} - \Delta_p \cdot \mathbf{p}^{*\prime T}_{pump}$, $\mathbf{p}_{UB} = \mathbf{p}^{*\prime}_{pump} + \Delta_p \cdot \mathbf{p}^{*\prime T}_{pump}$

3: **Generate population**: $\mathbf{x}_i = (x_i^1, x_i^2, ..., x_i^n), i = 1, ..., N_p$, $\mathbf{p}_{UB} < \mathbf{x}_i < \mathbf{p}_{LB}$

4: **Direct model**: Calculate $\mathbf{P}(f,z|\mathbf{x}_i) = \mathcal{R}(\mathbf{x}_i)$

5: **Calculate individual cost**: $J(\mathbf{x}_i)$

6: **While** $Ev < MaxEv$:

7:     **for** $i = 1 : N_p$ **do**

8:         **Select** three random individuals x_{r1}, x_{r2}, x_{r3} from the population where $i \neq r_1 \neq r_2 \neq r_3$

9:         **Mutation** : Form the donor vector as in Eq. 2

10:         **Crossover**: Form the trial vector as in Eq. 3

11:         **Check constraints**: Check if $\mathbf{p}_{LB} < \mathbf{u}_i < \mathbf{p}_{UB}$, otherwise, go to the next individual.

12:         **Direct model**: $\mathbf{P}(f,z|\mathbf{u}_i) = \mathcal{R}(\mathbf{u}_i)$

13:         **Evaluate**: if $J(\mathbf{u}_i) \leq J(\mathbf{x}_i)$, replace $\mathbf{x}_i$ with $\mathbf{u}_i$

14:         $Ev \leftarrow Ev + 1$

15:     **End for**

16: **End while**

III. SIMULATION RESULTS

We test the ability of the proposed framework to provide accurate pump power predictions for 2D signal power evolution profiles. The Raman solver defined in Eq. 1 is provided by GNPy library [30]. This function is utilized for the generation of the data-set for training, testing and validation of the CNN and can be also tuned to emulate a different amplification setup. Moreover, it is used in the cost function calculation process of the CNN-assisted DE framework illustrated in Algorithm 1.

We perform the power evolution design procedure jointly in spectrum and spatial domains for a single span of a single mode-fiber (SMF). The proposed pumping scheme is based on 8 pumps, including four counter- and four co- propagating pumps. In order to enable the framework to approach 2D power profiles of practical interest over the whole C-band, each set of co- and counter-propagating pumps consists of one second-order pump with the wavelength of 1366 nm and three first-order pumps with the wavelengths of 1425 nm, 1455 nm and 1475 nm, respectively. The pump wavelengths 1366 nm and 1455 nm are selected mainly based on the previously presented papers in which second-order and first-order pumps are utilized to achieve flat and symmetric power profiles [12], [13], [18]. Regarding the selection of the 1425 nm and the 1475 nm wavelengths, the main purpose has been to provide sufficient power for all the channels in the entire C-band [21], [31]. It is also worth mentioning that the pump wavelengths can be considered as the variables to be optimized, potentially

improving the performance [22]. Nonetheless, in the proposed framework, we have assumed fixed pump wavelengths to make the setup more similar to practical scenarios. Ranges of the pump powers for generation of the data-set are presented in table I.

TABLE I
RAMAN PUMP WAVELENGTHS AND POWER RANGES

Co-pumps	p_1	p_2	p_3	p_4
wavelength [nm]	1366	1425	1455	1475
power range [mW]	[200-1200]	[5-150]	[5-150]	[5-150]

Counter-pumps	p_5	p_6	p_7	p_8
wavelength [nm]	1366	1425	1455	1475
power range [mW]	[200-1200]	[5-150]	[5-150]	[5-150]

The 2D signal power profiles are analyzed over the whole C-band (between 192 THz and 196 THz) divided into 40 channels each with a bandwidth of 100 GHz. Input signal power per channel is set to 0 dBm resulting in the total signal power of 16 dBm. Moreover, the utilized SMF has the following specifications: length of the span $L_{span} = 80$ km, signal data attenuation of $\alpha_s = 0.2$ dB/km, second-order pumps $[p_1, p_5]$ attenuation $\alpha_P^{2nd} = 0.32$ dB/km, and first-order pumps $[p_2, p_3, p_4, p_6, p_7, p_8]$ attenuation $\alpha_P^{1st} = 0.25$ dB/km, effective area $A_{eff} = 80$ μm^2, non-linear coefficient $\gamma = 1.26$ 1/W/km, and Raman coefficient $g_R = 0.4125$ 1/W/km. Standard silica Raman efficiency profile is assumed [1].

The GnPy library [30] employs a BVP solver based on residual error control to solve the set of Raman differential equations. Considering this, we set the distance resolution to $z_{res} = 500$ m which satisfies the residual error threshold $r_{error} = 10^{-6}$. Since the selected distance step size is lower than the 1 km value reported in [22], the resulting 2D profiles have a larger dimension size compared to the ones in [22]. To have extra dimension reduction, we use four CNN layers rather than three layers proposed in the main CNN architecture. The other structural parameters of the network such as the number of the filters and their kernel size, sizes of the pooling layers and the number of nodes for the hidden layers are the same as reported in [22].

A. Evaluation of the CNN on achievable profiles

In order to train the CNN model, first, it is required to split the generated data-set into training, testing and validation sets. To realize the proper size of training data, training sets with different sizes from 1000 to 5000 samples are investigated and a validation set with 800 samples is used to evaluate the accuracy achieved with these training sizes. Validation MSE for different training sizes is shown in Fig. 3. Since there is not a considerable improvement in validation MSE error for training data-sets with more than 3500 samples, we choose this value as the training data size for all further analyses.

Once the CNN model is trained, its performance accuracy is measured using a test data-set with 800 samples. To measure the accuracy, we apply regression metrics based on the predicted and the true values of the pump setup for test data inputs. We choose coefficient of determination referred to as R-Squared (R^2) [32] as the measurement tool to interpret the
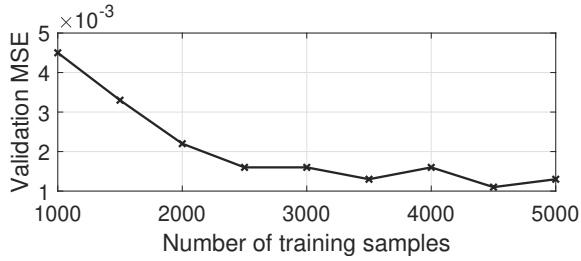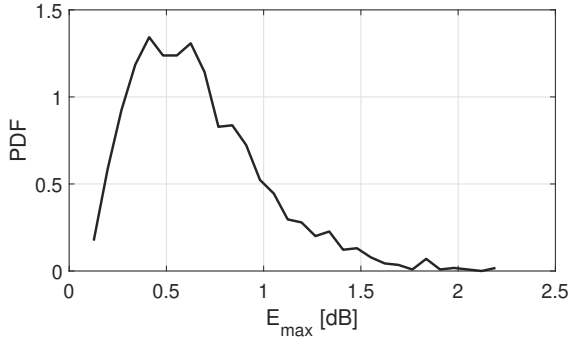
Fig. 3. Validation MSE as a function of number of training samples.

Fig. 4. PDF of E_{max} for the test data.

goodness of fit of the trained network. R^2 takes values between 0 and 1 where a R^2 of 1 indicates that the variation of the regression predictions perfectly explain the variation of the true values. Contrarily, low values of R^2 indicate that the regression model outputs do not vary according to the true values. R^2 values for predicted pump powers in test data-set are shown in table II. It is indicated that R^2 values are higher than 0.90 for all pumps except p_8 with R^2 slightly lower than 0.90. These results confirm the very good performance of the CNN in modelling the mapping between the 2D profiles and their corresponding pump power values in the proposed second-order amplification scheme.

TABLE II
R^2 VALUES OF THE PUMP POWER SET FOR TEST DATA

Pump	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
R^2	0.98	0.93	0.99	0.97	0.95	0.96	0.95	0.86

An alternative way to quantify the goodness of fit for the trained CNN model is to use the pump powers values predicted for the ith test sample as the input to the Raman solver in Eq. 1 to generate the corresponding 2D power profile $\hat{\mathbf{P}}^i(f, z)$. The error is defined as the maximum absolute difference E_{max}^i between the true 2D profile of ith test sample $\mathbf{P}^i(f, z)$ and the predicted 2D profile $\hat{\mathbf{P}}^i(f, z)$, both defined in dBm scale, as the following:

$$E_{max}^i = \max_{z,f} |\mathbf{P}^i(f, z) - \hat{\mathbf{P}}^i(f, z)| \quad (5)$$

Probability density function (PDF) of the E_{max} for test data-set is shown in Fig. 4. The low mean $\mu = 0.62$ dB and the standard deviation $\sigma = 0.33$ dB values for E_{max}^i also assert

the good performance of the proposed CNN model for the pump powers values prediction.

B. Evaluation of the CNN-assisted DE framework on designing specific profiles

Signal power profiles defined over the spectral and spatial domains such as 2D flat and symmetric with respect to the middle point of the span, are the two mostly addressed profiles in the literature due to their attractive properties [9], [16], [18], [23]. The CNN-assisted DE framework, as discussed in section II, aims to minimize a pre-defined cost value J. Mainly, the value of J depends on some requirements on the shape of the target 2D power profile. In the following, we explain the formulation of the cost function for 2D flat and symmetric power profiles. Once the cost function is defined, we present the results of applying the CNN-assisted DE technique discussed in the previous section.

1) 2D flat power evolution

We follow the formulation presented in the literature mainly for single channel and extend it to our wide-band with multi-channel (whole C-band) analysis [11], [23]. Additionally, we formulate the objective functions as a function of the pump powers values $\mathbf{p}_{pump}$ to fit properly in the proposed optimization framework.

We approach the 2D flat profiles design problem with three cost functions. Each of them applies a specific assessment quantity on the shape of the 2D profile. Considering this relation, the overall optimization problem is to find $\mathbf{p}_{pump}^*$ which jointly minimizes the three costs defined consecutively.

Assuming the signal power level defined in dBm scale, the main objective in a 2D flat profile design is to minimize the maximum power excursion in frequency and along the fiber span J_0 (in dB scale). Maximum power excursion is defined as the difference between the maximum and the minimum of the signal power level in spectral-spatial plane:

$$J_0(\mathbf{p}_{pump}) =$$
$$\max_{f,z}(\mathbf{P}(f, z|\mathbf{p}_{pump})) - \min_{f,z}(\mathbf{P}(f, z|\mathbf{p}_{pump})) \quad (6)$$

As detailed in section I, one of the aspirations in a wide-band amplification system is to achieve signal power with minimum spectrum excursion at the receiver side [4], [6], [7]. In this paper, we generalize its definition by aiming to minimize the maximum spectrum excursion among all distance points in the fiber. We refer to this cost function as spectrum excursion J_1 defined as the following:

$$J_1(\mathbf{p}_{pump}) =$$
$$\max_z[\max_f(\mathbf{P}(f, z|\mathbf{p}_{pump})) - \min_f(\mathbf{P}(f, z|\mathbf{p}_{pump}))] \quad (7)$$

Additionally, in the experimental setups presented in the literature for flat power evolution design [11], [23], the pump powers values are tuned empirically to provide a 0 dB gain between the transmitter and the receiver. The 0 dB gain J_2 between the transmitter and the receiver is considered here as the third cost value with the following formulation:

$$J_2(\mathbf{p}_{pump}) =$$
$$\max_f |\mathbf{P}(f, z = L|\mathbf{p}_{pump}) - \mathbf{P}(f, z = 0|\mathbf{p}_{pump})| \quad (8)$$

where L is the fiber length and $|.|$ is the absolute value operator.

Considering these objectives, we define a multi-objective optimization problem for 2D flat target profile aiming to find the best $\mathbf{p}_{pump}$ set. Due to the complexity of each cost function, it is challenging to realize whether the minimum value of the three costs occur at the same time or not. Therefore, to make the optimization process simpler and also to be able to control the impact of each objective, we make an approximation by converting it into a classical weighted-sum optimization. Approaching this, each objective is multiplied by a weight, defined as a hyper-parameter, and it is added to the other objectives as the following:

$$\mathbf{p}_{pump}^* = \arg\min_{\mathbf{p}_{pump}} \sum_{i=0}^{2} m_i J_i(\mathbf{p}_{pump}) = \mathbf{m} \cdot \mathbf{J}^T(\mathbf{p}_{pump}), \quad (9)$$

such that $\mathbf{p}_{LB} \leq \mathbf{p}_{pump} \leq \mathbf{p}_{UB}$, $m_i > 0$, $\sum_{i=0}^{2} m_i = 1$

where $\mathbf{m} = [m_0, m_1, m_2]$ is a hyper-parameter vector of weights, and each element of it controls the impact of the corresponding cost in the objective vector $\mathbf{J}(\mathbf{p}_{pump}) = [J_0(\mathbf{p}_{pump}), J_1(\mathbf{p}_{pump}), J_2(\mathbf{p}_{pump})]$, and T is the transpose operator. An advantage of the DE as a gradient-free optimization technique is that we can target problems with multiple objectives which we cannot solve easily with an ML-based inverse model like a CNN.

After defining the cost function for the 2D flat profile, we apply the framework presented in Fig. 2 to find the corresponding $\mathbf{p}_{pump}^*$. In this approach, first, we create a 2D power profile with 0 dBm power level as the 2D flat input to the CNN network to evaluate its performance in finding the corresponding set of pump powers values. The predicted pump powers by the CNN model for the 2D flat input profile are reported in table III, in the column labeled with CNN.

Following the framework in Fig. 2, the pump power values predicted by the CNN are used to initialize the fine-tuning process performed by the DE algorithm. Meanwhile, as presented in algorithm 1, DE has a set of hyper-parameters, which needs to be specified. A possible way to find the best set of hyper-parameters is to do a grid-search by setting the combination of all possible values of them. However, grid-search on finding the best set of hyper-parameters will make the problem complex and time consuming. In order to reduce the complexity, we set some of the DE hyper-parameters values such as the crossover probability and the mutation factor, same as the standard values $CR = 0.5$ and $F = 0.8$, normally used for other applications [26]. The population size N_p is set to 30 which is three times more than the dimension of the search space, recommended as a plausible value in the DE literature [26]. Additionally, the maximum number of cost function evaluation is set as $MaxEv = 1000$.

Regarding $\mathbf{\Delta}_p$, we set the same and fixed scaling factor value for the first-order pumps, and also the same value for the second-order pumps. Normally, second order pumps have

relatively high value and due to Raman pump-to-pump interaction, they have remarkable impact on the first order pump. Therefore, in $\mathbf{\Delta}_p$, we consider lower corresponding scaling factor for the second-order pumps compared to the first-order ones. Considering the higher predicted value of p_5 in comparison with the value predicted for p_1, we consider 1.4 [W] as the maximum possible value and consequently $\Delta_1 = \Delta_5 = 0.35$. Addressing the first-order pumps, based on the highest value which is predicted for p_3, we consider more deviation to first-order pumps by setting the corresponding scaling factor value to 0.5 resulting in $\mathbf{\Delta}_p = [0.35, 0.5, 0.5, 0.5, 0.35, 0.5, 0.5, 0.5]$ (0.35 scaling value for the second-order pumps and 0.5 for the first-order pumps).

TABLE III
PREDICTED PUMP POWER VALUES BY THE CNN AND CNN-ASSISTED DE WITH DIFFERENT WEIGHTS FOR 2D FLAT INPUT POWER PROFILE

model / Pump	CNN only	CNN+DE with $\mathbf{m}^{(1)}$	CNN+DE with $\mathbf{m}^{(2)}$	CNN+DE with $\mathbf{m}^{(3)}$
$p_1[mW]$	330	430	440	450
$p_2[mW]$	33	45	49	47
$p_3[mW]$	145	98	90	76
$p_4[mW]$	12	12	13	14
$p_5[mW]$	1030	1150	1060	990
$p_6[mW]$	12	8	6	6
$p_7[mW]$	19	12	18	21
$p_8[mW]$	43	24	63	63

According to Eq. 9, impact of the each cost $J_i, i = 0, 1, 2$ on the overall optimization problem is controlled by the corresponding weight value m_i. This approach provides the flexibility to target desired objectives with different weights. We consider the minimum power excursion defined in Eq. 6 as the objective with the highest impact, and accordingly, propose three scenarios with different values of $\mathbf{m} = [m_0, m_1, m_2]$. First, we set $\mathbf{m}$ as $\mathbf{m}^{(1)} = [1, 0, 0]$ where the objective is only the minimization of the power excursion J_0, secondly $\mathbf{m}^{(2)} = [2/3, 1/3, 0]$ where the objective is to minimize both the J_0 and the spectrum excursion J_1 with considering half impact than J_0. Finally, we set $\mathbf{m}^{(3)} = [2/3, 1/6, 1/6]$ where the objective is to minimize J_0 , J_1 and 0 dB gain variation J_2 giving equal impact to J_1 and J_2, each one with quarter impact of J_0. Table III shows the pump power values achieved for the CNN-assisted DE scenarios with the aforementioned $\mathbf{m}$ values.

To have an insight on the performance of the proposed 2D flat power profile design problem, we generate the power evolution profiles resulted from the CNN and also the CNN-assisted DE framework for different $\mathbf{m}$ values. Regarding this, the power evolution profiles over the distance consisting of 40 channels for the proposed scenarios are shown in Fig. 5.

Fig. 5 (a) shows the power evolution for the pump parameters predicted by the CNN, while Fig. 5 (b), (c) and (d) show the CNN-assisted DE optimization for $\mathbf{m}^{(1)}$, $\mathbf{m}^{(2)}$ and $\mathbf{m}^{(3)}$ values, respectively. The cost functions values J_0, J_1 and J_2 are specified on each one of the sub-figures and the corresponding values are reported in table IV. It is shown that the minimum power excursion is achieved for the CNN-
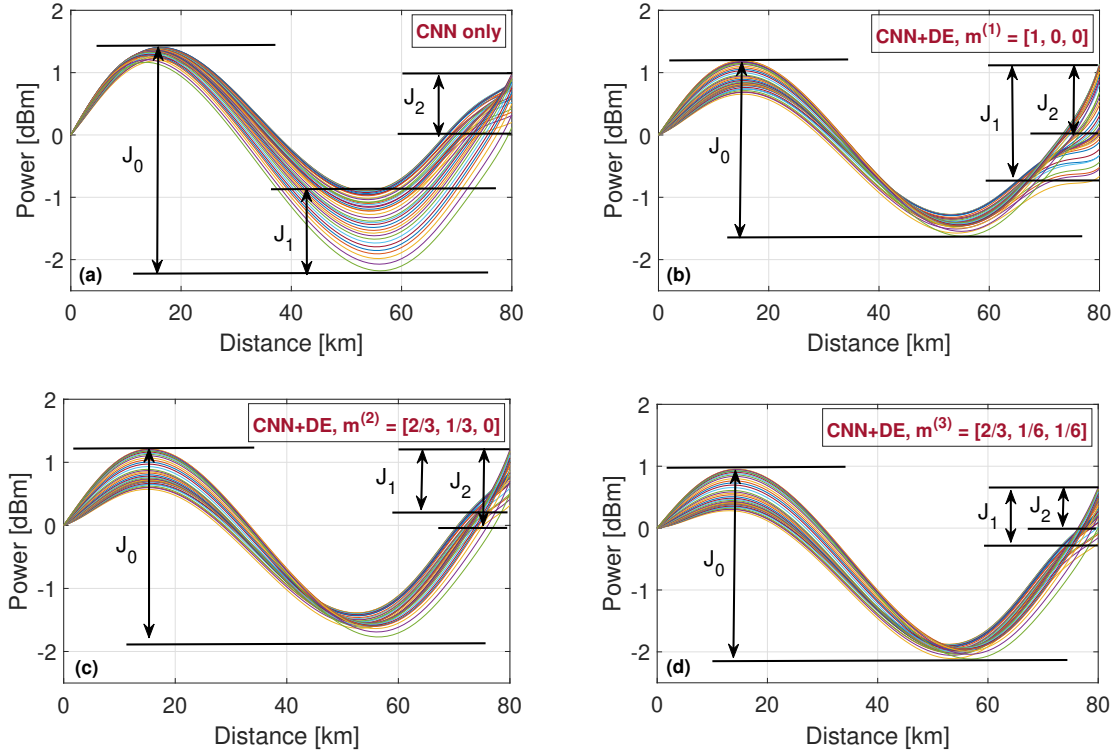
Fig. 5. Resulting 2D power evolution profiles for a 2D flat target profile with their corresponding objectives $[J_0, J_1, J_2]$. (a) CNN without the fine-tuning process, (b) CNN-assisted DE with $\mathbf{m}^{(1)} = [1, 0, 0]$, (c) CNN-assisted DE with $\mathbf{m}^{(2)} = [2/3, 1/3, 0]$, (d) CNN-assisted DE with $\mathbf{m}^{(3)} = [2/3, 1/6, 1/6]$.

assisted DE with $\mathbf{m}^{(1)}$ value which is 2.81 dB. However, this case has relatively high spectrum excursion J_1 and also high 0 dB gain J_2 variation values.

Considering the same framework with $\mathbf{m}^{(2)}$, the spectrum excursion improves almost 0.9 dB while the power excursion increases less than 0.2 dB. On the other hand, the 0 dB gain variation for both cases is still higher than 1 dB. By involving 0 dB gain variation in the fine-tuning process by using $\mathbf{m}^{(3)}$, it improves more than 0.5 dB while we observe an increase in the power excursion value by less than 0.1 dB. It has been illustrated that the position of the maximum spectrum excursion J_1 occurs at the end of the fiber link except for the output of the CNN model which is found at about the distance of 55 km (Fig. 5 (a)).

TABLE IV
COST FUNCTION VALUES ACHIEVED BY CNN AND CNN-ASSISTED DE FRAMEWORK FOR A 2D FLAT INPUT PROFILE

model Cost	CNN only	CNN+DE with $\mathbf{m}^{(1)}$	CNN+DE with $\mathbf{m}^{(2)}$	CNN+DE with $\mathbf{m}^{(3)}$
J_0	3.58 dB	**2.81** dB	2.97 dB	3.06 dB
J_1	1.48 dB	1.80 dB	**0.88** dB	0.90 dB
J_2	0.97 dB	1.14 dB	1.2 dB	**0.65 dB**

In addition to the wide-band analysis of power excursion in a 2D profile, we also analyze the power excursion as a function of frequency. For the predicted $\mathbf{p}^*_{pump}$ in each proposed 2D profile design scenario, the difference between the maximum and the minimum value of signal power level at each frequency value over the all distances is shown in Fig. 6.
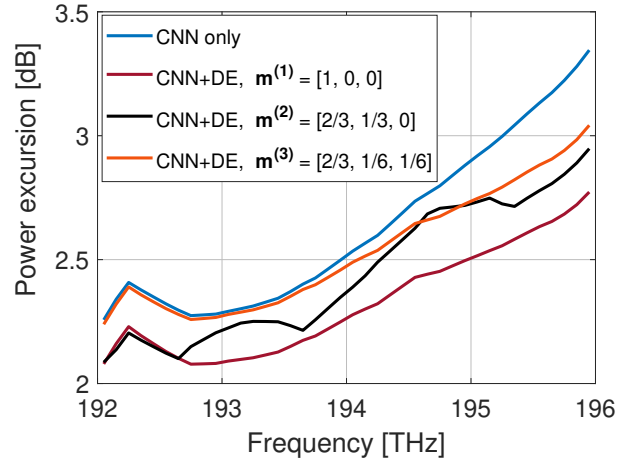

Fig. 6. Power excursion value computed as a function of frequency, shown for different pump power adjustment scenarios.

It is illustrated that generally, the power excursion value is increasing with frequency for CNN and also all the CNN-assisted DE scenarios. Additionally, the minimum power excursion which is less than 2.4 dB, is achieved almost at 192.7 THz. The maximum value for all scenarios is measured at 196 THz which is less than the corresponding power excursion J_0 measured based on Eq. 6 for the whole 2D frequency-distance plane.

In order to test the advantage provided by the CNN, we consider the case in which the optimizer is left free to span
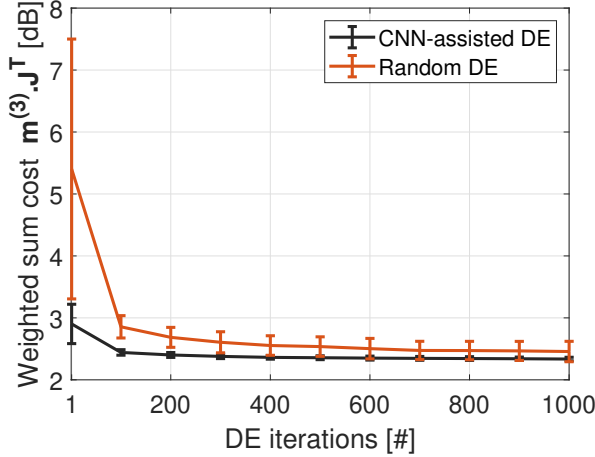
Fig. 7. Average cost (plot line) and standard deviation (error bars) for the CNN-assisted DE and the DE optimization with random initialization for flatness analysis as a function of number of DE iterations.

the whole pump power space, i.e. setting $\mathbf{p}_{LB}$ and $\mathbf{p}_{UB}$ to the minimum and the maximum pump power values specified in table I, respectively. In this case, referred to as "random DE", the DE population will be initialized in a wide space of pump power values. We run the algorithm for ten trials to see how is the average performance over the number of DE iterations for different randomly initialized populations. Additionally, we also perform ten trials for the CNN-assisted DE to compare its average performance with random DE. The objective of all trials for the both approaches is to solve Eq. 9 with $\mathbf{m}^{(3)}$.

In Fig. 7, the error bars of these two approaches are compared. The line plot determines the average error and the vertical error bar at each point shows one standard deviation above and below the average error. In average, the CNN-assisted DE converges to a lower minimum value of cost compared to the random DE. Furthermore, error bars show that the standard deviation of the cost value $\mathbf{m}^{(3)} \cdot \mathbf{J}^T$ at each iteration is relatively smaller for the CNN-assisted DE, indicating more reliability of using the CNN for initialization of the population by converging to the same minimum point for different trials.

2) 2D symmetric spatial power evolution

Symmetric power evolution in distance has been addressed so far mostly for single-channel narrow-band amplification scenarios [18], [33]. The main purpose for symmetric power evolution design is to numerically or experimentally minimize the asymmetry factor $A(f)$ at a desired frequency f_0 defined as the following [18]:

$$A(f_0) = \frac{\int_0^{L/2} |\mathbf{P}(f_0, z) - \mathbf{P}(f_0, L - z)| dz}{\int_0^{L/2} \mathbf{P}(f_0, z) dz} \qquad (10)$$

where the power profile $\mathbf{P}(f, z)$ is defined in linear scale [mW]. Generalizing the minimization model to a wide-band scenario, we define the proposed cost value J as the maximum asymmetry over all frequency channels existing in the bandwidth:

$$J = \max_f (A(f)) \qquad (11)$$

in which $A(f)$ and consequently J value, can be formulated as a function of the pump power values as the following:

$$J(\mathbf{p}_{pump}) =$$
$$\max_f \left[\frac{\int_0^{L/2} |\mathbf{P}(f_0, z | \mathbf{p}_{pump}) - \mathbf{P}(f_0, L - z | \mathbf{p}_{pump})| dz}{\int_0^{L/2} \mathbf{P}(f_0, z | \mathbf{p}_{pump}) dz} \right] \qquad (12)$$

and the optimization problem is formulated as:

$$\mathbf{p}_{pump}^* = \arg \min_{\mathbf{p}_{pump}} J(\mathbf{p}_{pump}) \qquad (13)$$

$$\text{such that} \quad \mathbf{p}_{LB} \leq \mathbf{p}_{pump} \leq \mathbf{p}_{UB}$$

Similar to the approach for flat power evolution design, after defining the cost value for this target profile, we apply the proposed CNN-assisted DE framework illustrated in Fig. 2.

The initial step in running the CNN-assisted diagram for designing a 2D symmetric power evolution is to provide a target 2D symmetric profile input for the CNN model. Unlike the flat power profile there are many 2D symmetric profiles which can be used as the input to the CNN model. In this paper, we have considered the second half-period of a sinusoidal signal as the symmetric power profile target for all channels. More specifically, we have formed a 2D symmetric power evolution, defined in logarithmic scale [dBm], with the following formulation:

$$\mathbf{P}^t(f, z) = 4sin(\pi z/L + \pi), \, \forall f, \, 0 < z < L \qquad (14)$$

This 2D power profile satisfies the symmetric condition in distance. Additionally, at each distance point, it has the same power level for all frequencies. Table V shows the set of pump power values achieved by both the CNN model and the CNN-assisted DE technique. It should be noted that all hyper-parameters required for the optimization have been set as we did for 2D flat power evolution design. Fig. 8 shows the results for 2D symmetric power profile design.

TABLE V
PREDICTED PUMP POWER VALUES BY CNN AND CNN-ASSISTED DE
OPTIMIZATION FOR SYMMETRIC POWER EVOLUTION PROFILE

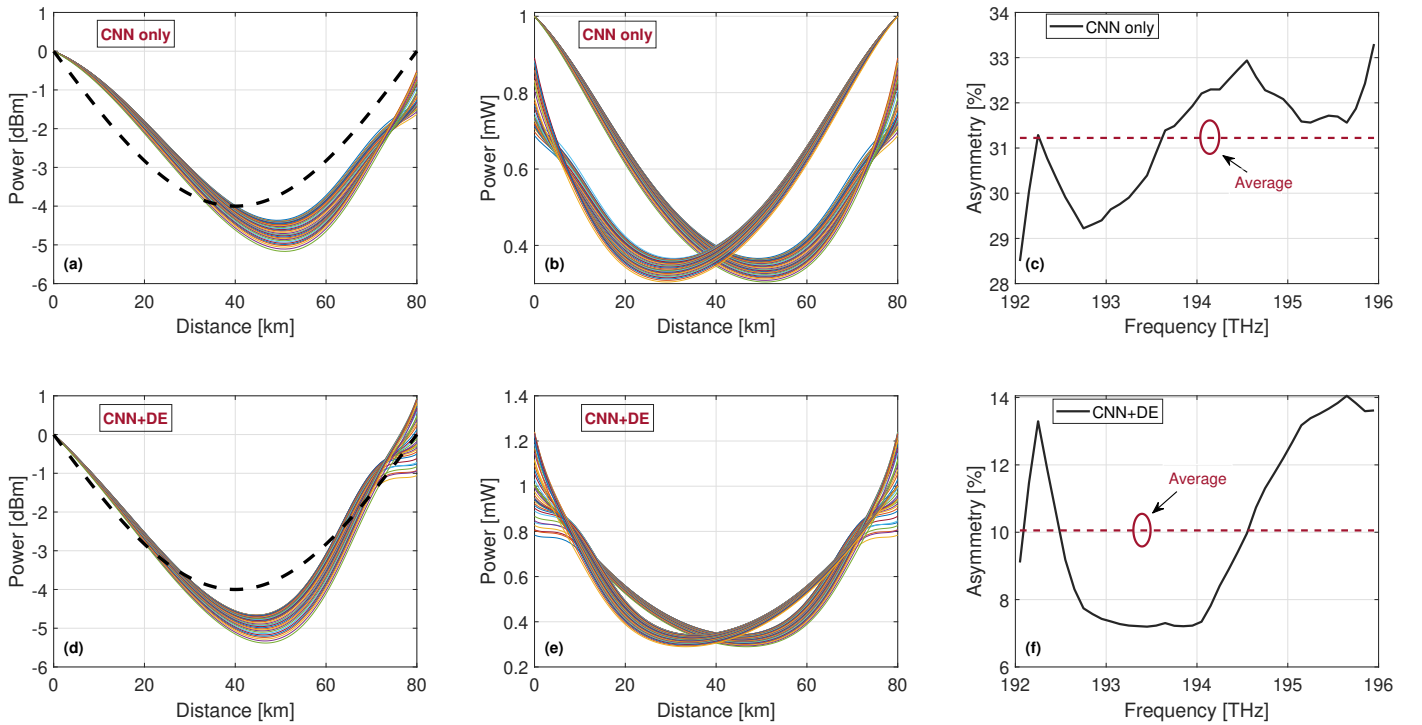| Pump \ model	CNN only	CNN-assisted DE
$p_1 [mW]$	150	100
$p_2 [mW]$	6	7
$p_3 [mW]$	63	36
$p_4 [mW]$	13	7
$p_5 [mW]$	1230	1480
$p_6 [mW]$	20	13
$p_7 [mW]$	10	6
$p_8 [mW]$	72	33

Fig. 8. Results of symmetric power evolution design. (a) The power evolution generated based on the pump power values predicted by the CNN model (dashed black curve is the symmetric target as a function of distance for all channels based on Eq. 14), (b) The power evolution and its reversed version in linear (mW) scale for the CNN, (c) The asymmetry in frequency for the CNN, (d) The power evolution achieved by CNN-assisted DE, (e) CNN-assisted DE power evolution results for symmetric input plotted with its reversed version in linear scale (mW) (f) The asymmetry in frequency for the CNN-assisted DE.

Fig. 8 (a) illustrates the power evolution resulting from the pump power values predicted for symmetric input with the dashed black curve showing the symmetric target profile generated based on Eq. 14. To have a visual understanding of the asymmetry minimization of power evolution defined in the linear scale [mW], we have shown the resulting 2D power evolution together with its reversed version in distance in Fig. 8 (b). Additionally, the asymmetry value has been shown as a function of frequency in Fig. 8 (c) asserting the maximum asymmetry of 33.2 % at 195.9 THz, and the minimum asymmetry of 28.5 % at 192.1 THz.

Furthermore, Fig. 8 (d) shows the power profile resulted from CNN-assisted DE with the dashed black curve indicating the symmetric target power evolution for all channels. Since the optimization process in Eq. 13 is performed to minimize asymmetry with power evolution defined in linear scale [mW] (Eq. 10), the resulting power evolution and the reversed version, both in linear [mW] scale, are shown in Fig. 8 (e). It is visually obvious that the asymmetry has improved strongly by applying DE optimization compared to its reversed version. Eventually, Fig. 8 (f) shows the asymmetry improvement for all frequencies with maximum value of 14% over the entire C-band, asserting more than 20% decrease in average asymmetry value by applying the fine-tuning process. Additionally, based on Fig. 8 (f), the resulting maximum 14% asymmetry is identified at 195.7 THz, and the minimum asymmetry is 7.2 %, which is identified at 193.8 THz.

Similar to the analysis performed for flatness, we consider

two pump power adjustment approaches to illustrate the impact of the CNN model used to initialize the DE population. In the first approach, the DE is assisted by the CNN model and in the second one, the CNN is not involved and the DE fine-tuning is performed by setting the $\mathbf{p}_{LB}$ and $\mathbf{p}_{UB}$ to the minimum and maximum values of pump powers, respectively.
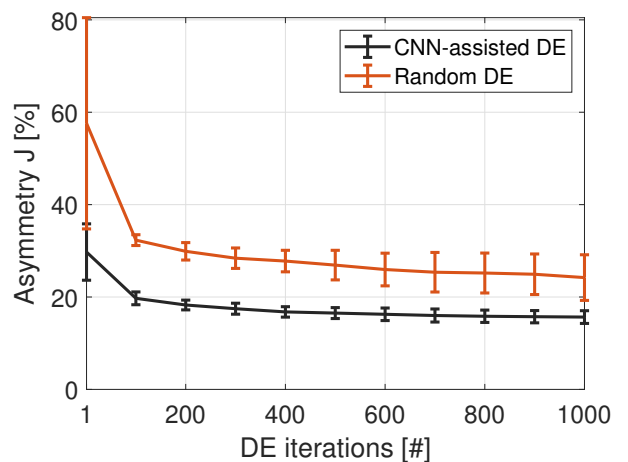

Fig. 9. Average asymmetry (marker) and standard deviation (error bars) for the CNN-assisted DE and the DE with random initialization as a function of number of DE iterations.

Fig.9 shows the line plot as the average error with the error bars for these cases indicating almost 10% average cost

improvement when CNN is involved in the population initialization. Moreover, it is obvious that the optimization without using CNN referred to as random DE has higher standard deviation at most of the points and this value increases with the number of evaluations. Hence, the probability of convergence to different minimum values is higher in case of running the DE algorithm blindly and without utilizing CNN to specify the initial solution.

IV. CONCLUSION

We proposed a CNN-assisted DE framework to adjust Raman pump power values for designing 2D target power profiles over the spectral and spatial domains in the fiber span. In the proposed framework, a pre-trained CNN model predicts the set of pump power values for a 2D target profile, and a DE algorithm performs gradient-free fine-tuning process aiming to improve the predicted values by the CNN. The CNN is trained on the data generated by exciting the amplification setup with randomly selected pump powers. However, its performance is not accurate when the target is to design specific 2D profiles of practical interest. To improve the pump powers predicted by the CNN, the DE employs a direct Raman solver function and performs fine-tuning to predict a new set of pump powers resulting in a feasible 2D profile with lower cost. The results assert the very good performance of the proposed CNN-assisted DE framework utilized for 2D flat and symmetric power profiles design. The proposed framework achieves power excursion of 2.81 dB for a flat target, and maximum asymmetry of 14% for a symmetric target in a 80 km fiber span in the whole C-band. Furthermore, the proposed DE with the CNN initialization provides higher accuracy with lower variance compared to the randomly initialized DE optimization.

ACKNOWLEDGMENT

REFERENCES

[1] C. Headley and G. P. Agrawal, *Raman Amplification in Fiber Optical Communication Systems*. Academic Press, 2005.

[2] W. S. Pelouch, "Raman amplification: An enabling technology for long-haul coherent transmission systems," *Journal of Lightwave Technology*, vol. 34, no. 1, pp. 6–19, 2016.

[3] U. C. de Moura, A. M. R. Brusin, A. Carena, D. Zibar, and F. D. Ros, "Simultaneous gain profile design and noise figure prediction for Raman amplifiers using machine learning," *Opt. Lett.*, vol. 46, pp. 1157–1160, Mar 2021.

[4] D. Zibar, A. M. Rosa Brusin, U. C. de Moura, F. Da Ros, V. Curri, and A. Carena, "Inverse system design using machine learning: The Raman amplifier case," *Journal of Lightwave Technology*, vol. 38, no. 4, pp. 736–753, 2020.

[5] J. Chen and H. Jiang, "Optimal design of gain-flattened Raman fiber amplifiers using a hybrid approach combining randomized neural networks and differential evolution algorithm," *IEEE Photonics Journal*, vol. 10, no. 2, pp. 1–15, 2018.

[6] U. C. de Moura *et al.*, "Multi–band programmable gain Raman amplifier," *J. Lightwave Technol.*, vol. 39, pp. 429–438, Jan 2020.

[7] G. Marcon, A. Galtarossa, L. Palmieri, and M. Santagiustina, "Model-aware deep learning method for Raman amplification in few-mode fibers," *Journal of Lightwave Technology*, vol. 39, no. 5, pp. 1371–1380, 2021.

[8] A. Carena, V. Curri, and P. Poggiolini, "On the optimization of hybrid Raman/erbium-doped fiber amplifiers," *IEEE Photonics Technology Letters*, vol. 13, no. 11, pp. 1170–1172, 2001.

[9] J. D. Ania-Castanon, "Quasi-lossless transmission using second-order Raman amplification and fibre Bragg gratings," *Opt. Express*, vol. 12, pp. 4372–4377, Sep 2004.

[10] J. Bouteiller, K. Brar, and C. Headley, "Quasi-constant signal power transmission," *2002 28TH European Conference on Optical Communication*, vol. 3, pp. 1–2, 2002.

[11] J. D. Ania-Castañón, V. Karalekas, P. Harper, and S. K. Turitsyn, "Simultaneous spatial and spectral transparency in ultralong fiber lasers," *Phys. Rev. Lett.*, vol. 101, p. 123903, Sep 2008.

[12] L. F. Mollenauer and K. Smith, "Demonstration of soliton transmission over more than 4000 km in fiber with loss periodically compensated by Raman gain," *Opt. Lett.*, vol. 13, pp. 675–677, Aug 1988.

[13] S. T. Le, J. E. Prilepsky, P. Rosa, J. D. Ania-Castañón, and S. K. Turitsyn, "Nonlinear inverse synthesis for optical links with distributed Raman amplification," *Journal of Lightwave Technology*, vol. 34, no. 8, pp. 1778–1786, 2015.

[14] S. Jansen, D. van den Borne, G. Khoe, H. de Waardt, P. Krummrich, and S. Spalter, "Phase conjugation for increased system robustness," in *2006 Optical Fiber Communication Conference and the National Fiber Optic Engineers Conference*, pp. 3 pp.–, 2006.

[15] I. Phillips *et al.*, "Exceeding the nonlinear-Shannon limit using Raman laser based amplification and optical phase conjugation," in *Optical Fiber Communication Conference*, p. M3C.1, Optical Society of America, 2014.

[16] M. Tan, M. A. Ai-Khateeb, M. A. Iqbal, and A. D. Ellis, "Distributed Raman amplification for combating optical nonlinearities in fibre transmission," in *2018 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*, pp. 1–2, IEEE, 2018.

[17] D. A. Chestnut, C. D. de Matos, P. Reeves-Hall, and J. Taylor, "Copropagating and counterpropagating pumps in second-order- pumped discrete fiber Raman amplifiers.," *Optics letters*, vol. 27 19, pp. 1708–10, 2002.

[18] P. Rosa, S. T. Le, G. Rizzelli, M. Tan, and J. D. A.-C. nón, "Signal power asymmetry optimisation for optical phase conjugation using Raman amplification," *Opt. Express*, vol. 23, pp. 31772–31778, Dec 2015.

[19] T. Ellingham, J. Ania-Castanon, R. Ibbotson, X. Chen, L. Zhang, and S. Turitsyn, "Quasi-lossless optical links for broad-band transmission and data processing," *IEEE Photonics Technology Letters*, vol. 18, no. 1, pp. 268–270, 2006.

[20] P. Rosa, G. Rizzelli, M. Tan, P. Harper, and J. D. Ania-Castañón, "Characterisation of random dfb raman laser amplifier for wdm transmission," *Opt. Express*, vol. 23, pp. 28634–28639, Nov 2015.

[21] A. E. Bednyakova, M. Fedoruk, P. Harper, and S. Turitsyn, "Hybrid gain-flattened and reduced power excursion scheme for distributed raman amplification," *Opt. Express*, vol. 21, pp. 29140–29144, Dec 2013.

[22] M. Soltani, F. Da Ros, A. Carena, and D. Zibar, "Inverse design of a Raman amplifier in frequency and distance domains using convolutional neural networks.," *Optics letters*, vol. 46 11, pp. 2650–2653, 2021.

[23] T. J. Ellingham, J. D. Ania-Castanon, R. Ibbotson, X. Chen, L. Zhang, and S. K. Turitsyn, "Quasi-lossless optical links for broad-band transmission and data processing," *IEEE Photonics Technology Letters*, vol. 18, no. 1, pp. 268–270, 2006.

[24] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[25] C. Blum, R. Chiong, M. Clerc, K. De Jong, Z. Michalewicz, F. Neri, and T. Weise, *Evolutionary Optimization*, pp. 1–29. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[26] S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 4–31, 2011.

[27] D. Whitley, "An overview of evolutionary algorithms: practical issues and common pitfalls," *Information and Software Technology*, vol. 43, no. 14, pp. 817–831, 2001.

[28] T. Eltaeib and A. Mahmood, "Differential evolution: A survey and analysis," *Applied Sciences*, vol. 8, no. 10, 2018.

[29] U. M. Ascher, R. M. M. Mattheij, and R. D. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*. Society for Industrial and Applied Mathematics, 1995.

[30] A. Ferrari *et al.*, "GNPy: an open source application for physical layer aware open optical networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 12, pp. C31–C40, 2020.

[31] J. Ania-Castanon, S. Kobtsev, A. Pustovskikh, and S. Turitsyn, "Simple design method for gain-flattened three-pump raman amplifiers," in *The 15th Annual Meeting of the IEEE Lasers and Electro-Optics Society*, vol. 2, pp. 500–501 vol.2, 2002.

[32] A. Colin Cameron and F. A. Windmeijer, "An r-squared measure of goodness of fit for some common nonlinear regression models," *Journal of Econometrics*, vol. 77, no. 2, pp. 329–342, 1997.

[33] P. Rosa, S. T. Le, G. Rizzelli, M. Tan, and J.-D. A.-C. nón, "Signal power asymmetry optimisation for optical phase conjugation using random DFB laser Raman amplification," in *Asia Communications and Photonics Conference 2015*, p. AM3D.5, Optical Society of America, 2015.