README.md

This dataset contains 51429 supercontinuum spectra generated by numerically simulating the propagation of short optical pulses in an optical fiber. The parameters characterizing both the input pulse and the optical fiber are provided below.

This dataset has been developed for and used in the research reported in the following publication:

> Diego Stucchi, Andrea Corsini, Goery Genty, Giacomo Boracchi, Alessandro Foi
> "A Weighted Loss Function to Predict Control Parameters for Supercontinuum Generation Via Neural Networks"
> 2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)

In this paper, we present a solution for solving the inverse problem of estimating generating parameters given an input spectrum, by means of a Fully-Connected Neural Network and a Convolutional Neural Network.

Simulation

The supercontinuum generation process was simulated by solving the Generalized Non-Linear Schrodinger Equation (GNLSE), including Raman scattering and the frequency-dependence of the nonlinear coefficient. The implementation consists of a standard split-step Fourier approach and a second order Runge-Kutta algorithm. Each generated spectrum is sampled over a different set of 4096 wavelength values $\Lambda \in \mathbb{R}^{4096}$, and consists of a vector of intensities $\sigma(\Lambda) \in \mathbb{R}^{4096}$.

Characteristics of the simulated optical fiber:

- Length (L): 6m
- Taylor-series dispersion coefficients at 1547 nm wavelength: $\beta_2 = -5.1 10^{-27} s^2/m$, $\beta_3 = 1.38 10^{-40} s^3/m$
- Nonlinear coefficient $\gamma = 2 10^{-3} W^{-1} m^{-1}$

Parameters of the input pulse:

- Central wavelength (λ_0);
- Duration expressed as Full Width at Half Maximum (FWHM);
- Peak power (Pp);
- Shape (hyperbolic-secant, fixed).

In the following table, we report the values used to model the simulated input pulse (wavelength, duration, peak power).

Parameter	Minimum	Maximum	Step
λ_0	1400nm	1700nm	10nm
FWHM	50fs	250fs	10fs
Pp	500W	20000W	250W

Folder

The dataset is organized in 3 `csv` files, each having 51429 rows.

- `parameters.csv` contains the triplets of generating parameters, in the following order: λ_0, FWHM, Pp;
- `wavelength.csv` contains the 4096 wavelengths where each spectrum is sampled;
- `spectra.csv` contains the spectra, expressed in arbitrary units (a.u.)

The triplet at row `r` in `parameters.csv` generates a the spectrum at row `r` of `spectra.csv`, whose wavelengths are stored at row `r` of `wavelength.csv`.

Usage

The database can be freely donwloaded and used for scientific research purposes. You are not allowed redistribute or modify it. Proper citing of this resource is expected if the database is used in research or other reporting.

The database is meant to be useful, but it is distributed WITHOUT ANY WARRANTY, and without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.

Inclusion of this database or even parts of it in a proprietary program is not allowed without a written permission from the owners of the copyright. If you wish to obtain such a permission, you should contact:

`diego.stucchi@polimi.it`