23ème Congrès de Chimiométrie, Université de Bretagne Occidentale, Brest, Juin 2022.

Deep learning based correction and data modeling techniques as tools for chemometricians

T. Bocklitz[1,2]

[1] Leibniz Institute of Photonic Technology (IPHT), Albert-Einstein-Straße 9, 07745 Jena, Germany

[2] Institute of Physical Chemistry and Abbe Center of Photonics (IPC), Friedrich-Schiller-University, Helmholtzweg 4, 07743 Jena, Germany

Keywords: Correction procedures, deep learning, data modeling.

1 Introduction

Chemometrics is the science focusing on the analysis of chemical data. Within chemometrics the analysis tasks are typically carried out using workflows or data pipelines, which combine correction procedures, preprocessing methods and data modeling techniques, e.g. classical machine learning. In the latter method the combination of feature extraction or selection with easy machine learning models is carried out. In the last decades another type of machine learning called deep learning (DL) emerged, which is typically based on (highly parametrized) artificial neural network geometries, which does an implicit feature extraction.

2 Theory

A large ‘zoo’ of deep learning models exists and can be applied to chemical, spectroscopic and image data [1]. The methods used in this contribution are convolutional neuronal networks (CNNs) Long short term memory networks (LSTMs) and autoencoder methods like the U-Net model. In a nutshell, a CNN is a feed-forward neuronal network which performs convolutional operations in parts of the neurons/layers, while the LSTM has feedback connections and can be used to process sequence data. The U-Net is an autoencoder structure with shortcut connection. Beside these general types of deep learning models, generative adversarial networks (GANs) are utilized in this contribution, which are combinations of generator and discriminator networks.

3 Material and methods

In this contribution, IR spectra of Polymethylmethacrylat (PMMA) objects and simulation of these samples are utilized [2]. In reference [3] CARS spectra of solvents are measured, while in reference [4] Raman spectra of inflammatory bowel disease tissue sections (IBD) are analyzed. In the study [5] multimodal images of colon tissue section are used for modality transfer.

4 Results and discussion

In this overview the usage of deep learning techniques for chemical data such as spectra and images are demonstrated. In [2,3] the advanced correction of measurement artifacts or more general side contributions of the measurement using DL methods is demonstrated. Specifically, based on simulations, artifacts introduced by the shape of measured objects could be corrected within IR spectra using a 1D U-Net architecture [3]. In reference [3] an extraction of the imaginary part of the non-linear susceptibility from the measured square modules, e.g. CARS spectra, using the LSTM network is demonstrated. Beside these correction based usage of deep learning, these DL techniques can be used for data modeling as well [4,5]. While in reference [4] a CNN is adapted to Raman

spectra for tissue diagnostics, GANs and cycle GANs are used in the study [5] to transfer multimodal images to another measurement modality.

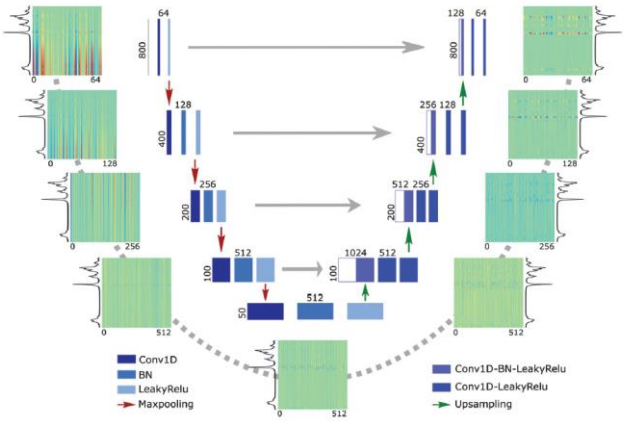

Figure 1 – Visualization of a U-Net structure to correct IR spectra (image retrieved from [2]).

5 Conclusion

Deep learning techniques as highly parametrized models feature a unique potential to solve problematic tasks in data modeling and data correction. These DL models can be used as tools for the chemometrician if large datasets with independent measurements are available or simulations can be done reliably.

6 References

[1] Pradhan, P.; Guo, S.; Ryabchykov, O.; Popp, J. & Bocklitz, T. W. Deep learning a boon for Biophotonics? *Journal of Biophotonics*, 13, e201960186, 2020

[2] Guo, S.; Mayerhöfer, T.; Pahlow, S.; Hübner, U.; Popp, J. & Bocklitz, T. Deep learning for 'artefact' removal in infrared spectroscopy *Analyst*, 145, 5213-5220, 2020.

[3] Houhou, R.; Barman, P.; Schmitt, M.; Meyer, T.; Popp, J. & Bocklitz, T. Deep learning as phase retrieval tool for CARS spectra *Optics Express*, 28, 21002-21024, 2020.

[4] Kirchberger-Tolstik, T.; Pradhan, P.; Vieth, M.; Grunert, P.; Popp, J.; Bocklitz, T. W. & Stallmach, A. Towards an interpretable classifier for characterization of endoscopic Mayo scores in ulcerative colitis using Raman Spectroscopy *Analytical Chemistry*, 92, 13776-13784, 2020.

[5] Pradhan, P.; Meyer, T.; Vieth, M.; Stallmach, A.; Waldner, M.; Schmitt, M.; Popp, J. & Bocklitz, T. Computational tissue staining of non-linear multimodal imaging using supervised and unsupervised deep learning *Biomedical Optics Express*, 12, 2280-2298, 2021