 https://doi.org/10.1051/epjconf/202430905002

A convolutional neural network approach for multilayer analysis in infrared nanospectroscopy

Dario Siebenkotten and *Bernd* Kästner*

Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin, Germany

Abstract. The combination of Fourier-transform infrared spectroscopy and scattering-type scanning near-field optical microscopy allows for the spectroscopic investigation of materials and structures at the nanoscale, far below the diffraction limit in the infrared. This resolution is achieved by the use of metallized atomic force microscopy tips which locally illuminate the sample through the creation of near-fields at their apex. The complex interaction between incident light, a metallized tip and a layered sample necessitates the use of sophisticated models. While these models are powerful, using them to fit measured spectra is generally slow and often unstable, thus requiring expert oversight. Neural networks present a fast and often more stable alternative, but their application so far has focused on bulk samples. Here, we present the use of convolutional neural networks for the recovery of the optical and thickness properties from the spectra of samples consisting of one or two layers of polar crystals on silicon.

1 Nanoscale infrared spectroscopy

Infrared imaging and spectroscopy provide access to the plasmon response of many doped semiconductors, phonon bands in many materials and typical molecular rotational and vibrational excitations. However, the diffraction limit poses heavy restrictions on the sample structures that can be investigated with most methods in the infrared. The need to circumvent the diffraction limit thus gave rise to the development of atomic force microscopy (AFM) based methods in the infrared, the most widely used such technique being scattering-type scanning near-field optical microscopy (s-SNOM). In s-SNOM, the incident light is focused on a metallized AFM tip which leads to the creation of three-dimensionally highly confined near-fields at its apex. These near-fields are reflected by the sample directly underneath the tip in dependence of its optical properties and then scattered into far-field by the tip, where they are measured. A schematic of the measurement principle is shown in Fig. 1, including the exponentially decaying near-fields penetrating into the sample. As the lateral extension of the near-fields is significantly smaller than that of the diffraction-limited incident light spot, the signal is dominated by background contributions. To extract the near-field contribution, the tip is oscillated and the resulting, near-field-dominated, higher harmonics are measured via lock-in detection. With increasing harmonic order the near-field ($S_2 - S_4$) gets more tightly confined, as indicated by the different colours in Fig. 1.

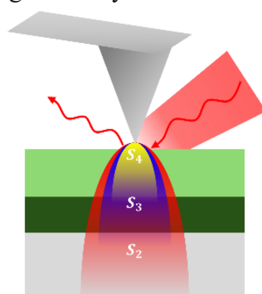

Fig. 1. Schematic of the measurement principle of s-SNOM.

For many samples, the properties of interest are encoded in its local spectral response. To access this information, s-SNOM is combined with a broadband light-source and an asymmetric Michelson-interferometer (nano-FTIR). The prediction of nano-FTIR spectra, particularly for layered samples, requires sophisticated models, such as the multilayer finite dipole model [1]. While such models generally solve the forward problem, i.e., the prediction of a measurand from known physical properties, they do not inherently solve the inverse problem, i.e., the recovery of physical properties from measurands. While models have been proposed for the specific cases of bulk [2] and non-resonant samples [3] with a single thin layer, the most common approach to the solution of the inverse problem is the combination of the forward model with a fitting algorithm. Practically, the fits with these models are generally slow and often prove unstable, requiring oversight by a human expert. The application of neural networks to the inverse problem presents a suitable alternative, as they, after training, improve the evaluation speed massively and tend to also offer increased stability. Neural networks have been shown to produce good results for the solution of the inverse problem on bulk samples on purely simulated [4,5] and experimental [6] data sets. Here, we present convolutional neural networks trained with simulated spectra to solve the inverse problem for a layer stack

* Corresponding author: Bernd.Kaestner@ptb.de

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

comprised of a silicon substrate and one or two polar crystal layers, which are of specific interest as they support resonant coupling between the tip and surface phonon polaritons.

2 Results

We choose a one-dimensional convolutional neural network (CNN) architecture with an input consisting of three harmonics, the 2nd to 4th, as they are experimentally accessible and have different spatial extensions into the sample). The spectrum at each harmonic is split into amplitude and phase. A CNN architecture is chosen, as they excel at feature identification, and clear features in the spectra are expected due to the surface phonon polariton resonances. The polar crystal layers of interest are characterized by a single phonon band, which can be modelled by a single Lorentz oscillator, mathematically defined by

$$\varepsilon(\omega) = \varepsilon_\infty \left(1 + \frac{\omega_{LO}^2 - \omega_{TO}^2}{\omega_{TO}^2 - \omega^2 - i\omega\gamma}\right), \qquad (1)$$

with ω being the angular frequency. The material specific parameters of each layer are longitudinal and tangential angular phonon frequencies ω_{LO} and ω_{TO}, the damping rate γ and the high frequency relative permittivity ε_∞. These parameters are recovered for each layer simultaneously, in addition to the layer thickness. We choose the same network architecture, which is shown in Fig. 2, for the single layer and the double layer cases, which differ only in the number of output parameters. The

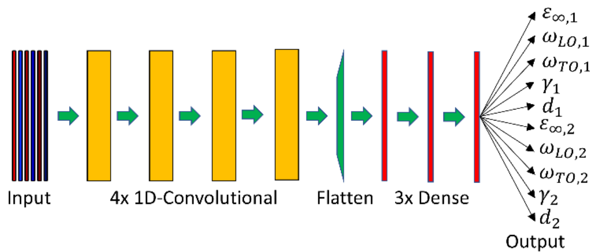

Fig. 2. The used convolutional neural network architecture. The output (here for the two layer case) is the only difference between the two cases.

network consists of four one-dimensional convolutional layers for feature extraction, and three dense layers for recovery of the layer parameters from the extracted features. For the single layer samples a data set of 100.000 spectra was generated using the multilayer finite dipole model. For the two-layer samples this number was increased to 500.000. In both cases 80% of the data sets were used for training and 20% for validation to prevent overfitting. An example of a typical input and the corresponding reconstructed spectrum for the two-layer case, using the recovered parameter and the multilayer finite dipole model, is shown in Fig. 3. In the spectrum two resonances are clearly visible, corresponding to the surface phonon polaritons of the two polar crystals. The resonance at 1900 cm^{-1} appears due to the interface to air of the upper layer. The second resonance originates at the

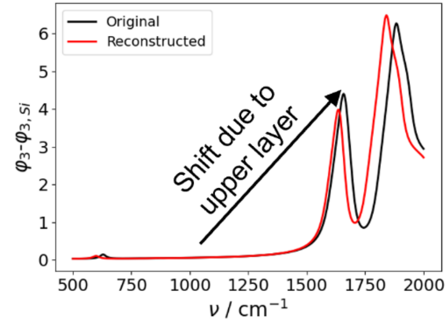

Fig. 3. Typical example of the reconstruction of a two-layer sample spectrum. The MAE of the recovered parameters is a little over average with 11.6%.

interface between the two layers and appears significantly shifted compared to the expectation if there was no upper layer (marked by the arrow). Such spectral shifts complicate the parameter recovery, but also highlight why sophisticated models are needed for the recovery.

To give an intuitive quantification of the performance of the data, the normalized medium absolute deviation (MAE) between the recovered and ground-truth parameters, $\left|(p_{extr} - p_{orig})/p_{orig}\right|$, is shown in Table 1. for the different parameters and both cases. The recovery of the frequency parameters performs very well, while the high frequency permittivity and thickness perform notably worse, particularly for the buried layer in the two-layer case. This is not unexpected, given the exponential decay of the near-fields into the material, leading to the second layer often contributing only weakly to the total scattered signal, particularly for larger thicknesses. The average MAE is below 5% in the one-layer and below 10% in the two-layer case.

Table 1. Median absolute error in percent for the different recovered parameters for the single and two-layer cases.

	$\varepsilon_{\infty 1}$	ω_{LO1}	ω_{TO1}	γ_1	d_1	$\varepsilon_{\infty 2}$	ω_{LO2}	ω_{TO2}	γ_2	d_2
One-Layer	3.5	0.8	1.4	2.3	5.8	-	-	-	-	-
Two-Layer	9.3	2.8	4.0	9.2	11.0	16.7	4.3	6.0	10.1	21.7

References

1. B. Hauer, A. P. Engelhardt, T. Taubner, Opt. Express. **20** (2012)
2. A. A. Govyadinov, I. Amenabar, F. Huth, P. S. Carney, R. Hillenbrand, J. Phys. Chem. Lett. **4** (2013)
3. A. A. Govyadinov, S. Mastel, F. Golmar, A. Chuvili, P. Scot Carney and R. Hillenbrand, ACS Nano **8** (2014)
4. X. Chen, R. Ren and M. Liu, Phys. Rev. Applied **15** (2021).
5. Y. Zhao, X.Chen, Z. Yao, M. Liu and M. M. Fogler, J. Appl. Phys. **7** (2023)
6. X.Chen, Z. Yao, S. Xu, A. S. McLeod, S. N. Gilbert Corder, Y. Zhao, et al., ACS Photonics **8** (2021)