Deconvolution of 1D NMR spectra: A deep learning-based approach

N. Schmid [a,b,*], S. Bruderer [d], F. Paruzzo [d], G. Fischetti [c], G. Toscano [d], D. Graf [d], M. Fey [d],
A. Henrici [a], V. Ziebart [a], B. Heitmann [d], H. Grabner [a], J.D. Wegner [b], R.K.O. Sigel [b], D. Wilhelm [a]

[a] *Zurich University of Applied Sciences (ZHAW), Switzerland*
[b] *University of Zurich (UZH), Switzerland*
[c] *Ca' Foscari University of Venice, Italy*
[d] *Bruker Switzerland AG, Switzerland*

ARTICLE INFO

ABSTRACT

The analysis of nuclear magnetic resonance (NMR) spectra to detect peaks and characterize their parameters, often referred to as deconvolution, is a crucial step in the quantification, elucidation, and verification of the structure of molecular systems. However, deconvolution of 1D NMR spectra is a challenge for both experts and machines. We propose a robust, expert-level quality deep learning-based deconvolution algorithm for 1D experimental NMR spectra. The algorithm is based on a neural network trained on synthetic spectra. Our customized pre-processing and labeling of the synthetic spectra enable the estimation of critical peak parameters. Furthermore, the neural network model transfers well to the experimental spectra and demonstrates low fitting errors and sparse peak lists in challenging scenarios such as crowded, high dynamic range, shoulder peak regions as well as broad peaks. We demonstrate in challenging spectra that the proposed algorithm is superior to expert results.

1. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a powerful and diverse tool for the quantitative analysis of molecules. Spectra can vary from a few peaks to highly overlapping regions with hundreds of peaks. The process of retrieving these peaks, i.e. signals, information from a 1D NMR spectrum is called deconvolution. Deconvolution is an essential step for the quantification, elucidation of the structure and verification of the structure of small molecules, and it facilitates workflows, for example, in drug discovery and the study of molecules [1]. In detail, the deconvolution process obtains a list of peaks and their parameters from a spectrum without additional information on the molecular structure.

Creating a spectrum from a peak list, known as convolution, is a straightforward and well-defined task if distortions are neglected. On the other hand, inverting this operation, that is, determining a peak list from a given spectrum, is an ill-posed inverse problem [2]. Hence, there is an infinite number of solutions to a deconvolution problem. A solution with a small residual error (the difference between the original and reconstructed spectrum) is not necessarily practical since too many peaks could be picked to fit the noise, distortions, or other influences. This notion is also known as over-fitting. To eliminate overfitting and decide which of all possible solutions to a deconvolution problem to choose, one needs to add information in the form of regularization or constraints. But even the commonly used approaches of sparsity promoting regularization meet their limits, depending on the regularization used [3,4]. Forcing the solution to be sparse can still lead to ambiguous results, see Fig. 1: In this example, experts usually prefer the sparser three-peak solution, although the spectrum was created using the five-peak annotation. Peaks are particularly challenging to be resolved and an expert could not distinguish between one or two peaks, especially if the chemical shift of two peaks is nearly identical, cf. Fig. 1 (bottom, center), or if the amplitudes are factors of magnitude apart, cf. Fig. 1 (bottom, right). Increasing the number of detected peaks, i.e. model order, and thus, the number of free parameters for the fitting, will generally reduce the mean absolute error (MAE), reflected by the lower residual error of the five-peak fit. Furthermore, the intrinsic ambiguity of the measurement process and inaccurate spectral processing steps contribute to further unknown physical and experimental factors. Together, these result in line shape distortions, spectral artifacts, and low signal-to-noise ratio (SNR). All of these factors make deconvolution of NMR spectra challenging for both experts and machines.

The deconvolution processes typically require significant manual intervention, operator supervision, and expert knowledge to achieve a good result. However, increasing the automatization of

---

* Corresponding author.
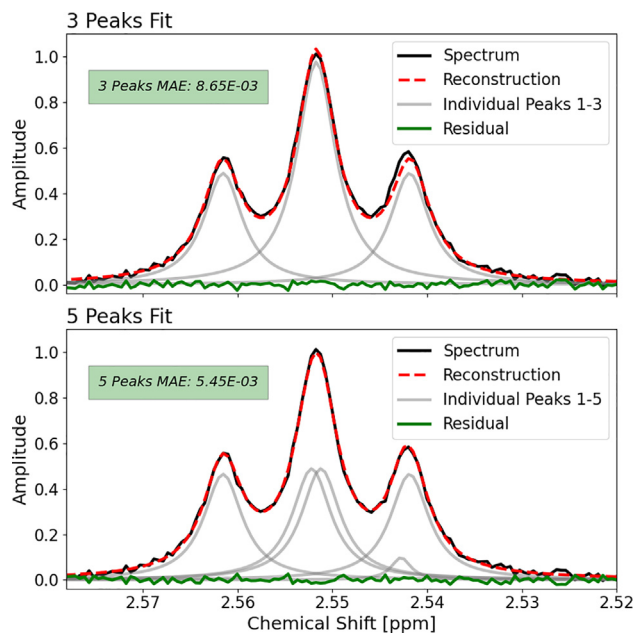 *E-mail address:* scdn@zhaw.ch (N. Schmid).

Fig. 1. Deconvolution is an ill-posed problem. The substantially different solution will produce almost the same MAE of the shown region. For example, the spectrum (black) can be reconstructed (red, dashed line) using three (grey, top) or five peaks (grey, bottom).

the deconvolution process can enhance the accessibility of NMR spectroscopy for non-experts, improve reproducibility, and reduce labor costs. Thus, automatic extraction of various peak parameters in NMR spectroscopy is essential to facilitate the NMR workflow.

Classical approaches focus on information based on simple signal properties such as integral, amplitude, SNR, or geometric features [5,6]. One of the first steps in deconvolution methods usually consists of determining the number of peaks and their corresponding positions. Various approaches such as wavelet [7], matrix factorization [8], singular value decomposition [9], and Bayesian methods [10] have been proposed for the peak picking problem. Classical approaches that cover the complete deconvolution workflow often rely on the fitting of the free induction decay (FID) or the spectrum in the frequency domain, combined with an explicit criterion for determining the number of peaks, i.e., model order selection. However, trading off fitting error against model order can be pretty cumbersome [3,4]. The so-called CRAFT (Complete Reduction to Amplitude-Frequency Table) algorithm is an example of a classical FID-based fitting approach [11]. In contrast, another whole family of approaches is based on the autoregression properties of the FID, for example, Hankel matrix [12], filter diagonalization [13], and fast padé transform-based methods [14]. They are based on the fact that the subsequent points in the FID signal can be modeled as a linear combination of the preceding ones. On the other hand, an approach for deconvolution in the frequency domain is the Global Spectral Deconvolution (GSD) algorithm from Cobas et al. [15], which relies on the spectrum and its first and second derivatives for peak picking and fitting. Unfortunately, many classical algorithms for deconvolution of 1D NMR spectra still need some expert knowledge to set the hyperparameters of the method, rely on simple and fixed patterns, such as the spectrum and its derivatives, and exhibit difficulties in finding the correct number of peaks, that is, in the selection of the model order [3]. Instead of using simple and fixed patterns such as the signal and its derivatives or simple geometrical features, machine learning techniques can be utilized to learn more expressive, domain-specific features for deconvolution, such as multiplet patterns. For example, triplets

are a typical pattern in NMR spectra that can be recognized; see Fig. 1.

Machine learning methods, particularly deep learning methods, have proven their powerful capabilities to find and exploit patterns in all types of data. They have become indispensable in many domains, such as natural language processing or computer vision, to extract expressive features from nonlinear, high-dimensional signals. Deep neural networks (DNNs) are suitable for a wide range of learning tasks, ranging from classification and regression to unsupervised learning.

Neural networks have a long history in NMR spectroscopy, with roots dating back to the 1980s [16]. Despite the early beginnings, most practical deep learning algorithms in NMR have only emerged in recent years due to improved algorithmic performance and computational capabilities [17]. Deep learning methods have been successfully applied in other areas of spectroscopy, for example, in IR spectroscopy to identify functional groups [18] or to process two-electron dipolar spectroscopy (DEER) data [19]. However, NMR spectroscopy also seems well suited for applying deep learning techniques, primarily because of its richness of information, quantitative nature, and high reproducibility. Chemical shift prediction is one of the most prominent applications [20–24]. Especially in the field of 2D NMR spectroscopy, deep learning has found various applications in recent years. The reconstruction of 2D spectra from nonuniformly sampled (i.e., undersampled) spectra (NUS) [25–27], or the denoising of low-SNR spectra [28] has been tackled. Deep learning-based peak picking in 2D-4D spectra (ARTINA) was carried out by Klukowski et al. [29]. In terms of learning-based deconvolution approaches, there are time and frequency domain techniques. Huang et al. [30] propose a deep Hankel matrix algorithm for the deconvolution of FID data. Recently, Li et al. [31] introduced a deconvolution method for 2D experimental NMR spectra that handles strongly overlapping, relatively narrow peaks called DEEP Picker. Generally, frequently recurring multiplet patterns and the complexity of inverse problems in NMR spectroscopy are factors in favor of the strengths of deep learning. This has led to the aforementioned promising applications of deep learning in spectroscopy, particularly in 2D NMR spectroscopy.

However, applications in the more common high dynamic range (HDR) 1D NMR spectroscopy and the demonstration of robust transferability to experimental 1D NMR spectra for DNNs trained on synthetic data are not fully solved [17]. Especially in challenging regions with strongly overlapping, very broad peaks, shoulder peaks, and HDR issues, many methods struggle to demonstrate robust, expert-level performance [32].

Our method addresses these problems: We propose a deep learning-based deconvolution algorithm for 1D NMR spectra with expert-level performance in various experimental spectra. More precisely, our contributions, which will be explained and discussed later in detail, are the following:

- Creation of realistic synthetic 1D NMR data set that allows a robust model transfer to experimental spectra
- An implicit regularization approach through efficient automatic labeling of these synthetic spectra
- Custom data pre-processing for high dynamic range regions and broad lines
- A neural network for the training with the synthetic spectra and the application to the experimental spectra that handles broad lines and their overlap with narrow lines, highly crowded regions, and HDR spectra
- Altogether, a fully automated method that yields expert-level quality sparse deconvolution results on experimental spectra

The paper is organized as follows: Section 2 describes the data creation, pre-processing, and labeling, as well as the deep learning

method. Section 3 presents the results of the deconvolution in synthetic and experimental spectra and compares them with expert annotations. Furthermore, different aspects of deep learning as a tool for deconvolution are discussed. Finally, Section 4 concludes. Various technical details of our setup are explained in the SI Section.

2. Methods

We present a combined deconvolution algorithm that uses deep learning and classical methods. The DNN is trained using a synthetic data set that has been pre-processed and labeled with an NMR data-specific approach; cf. Fig. 2 (top row). In a second step, the trained model is applied to the pre-processed experimental spectra, cf. Fig. 2 (bottom row), and yields, together with a subsequent automatic classical peak parameter fine-tuning, a peak list.

2.1. Data creation for robust model transfer from synthetic to experimental spectra

Consistently labeled and high-quality experimental NMR data are expensive and generally unavailable in the quantities needed to train modern neural networks [1]. Thus, our method is based on synthetic training data sets [17] that are freely available and can be generated in large amounts. Distortion- and artifact-free synthetic data can be generated from first principles, since NMR is a well-understood physical phenomenon. However, creating perfectly realistic synthetic data, including distortions and artifacts, is challenging, often leading to a slight discrepancy between synthetic training data and experimental test data. The notion of having different types of data sets for training and testing in supervised learning is known as distribution shift [33] and dealing with distinct distribution shifts is a common problem in machine learning called transfer learning [34]. We approach this problem from two sides.

Firstly, we make our training spectra as realistic as possible. For data creation, we use synthetic multiplet structures consisting of pseudo-Voigt lines [35]. Pseudo-Voigt lines correspond to an experimental Gaussian line shape distortion to the theoretical Lorentzian resonance profiles. Distortions, such as inaccurate phase and baseline corrections, are applied to the pure signal. We then add a small linear frequency-dependent phase error to the spectrum for phase distortions. Incorrect baseline correction is emulated with randomly sampled points connected through a spline smoothing model. These procedures serve as NMR spectroscopy-specific data augmentation, which helps to prevent overfitting and promote transferability to experimental spectra [36]. Further-

more, Gaussian noise is added to the spectrum to produce a random maximum SNR level in the range of $[10^2, 10^4]$, cf. SI 5.1 for details and an illustration of the data creation process. An advantage of creating a custom synthetic data set is that the data distribution is malleable to cover the space of all possible experimental spectra.

Secondly, we process the experimental spectra to match the training spectra. Therefore, we correct the experimental spectra with the automatic phase and baseline correction command "apbk" from TopSpin [37]. The corrected experimental spectra are then line-broadened with 0.3 Hz and resampled to match the number of points per Hz in the training spectra. This steps ensure that the peaks in the spectrum have approximately Voigt line-shapes with the same width range with respect to the number of points as in the training set. Altogether, the realistic data set creation and the experimental spectra adaption reduce transfer learning issues as much as possible.

2.2. Deep learning-specific data pre-processing

Peaks in 1D NMR spectra can have parameters that span orders of magnitude. Some peaks in a spectrum have a very low SNR, e.g., below 10; others have a very high SNR, e.g., above 10^4. Such a broad range of signal intensities is referred to as a high dynamic range (HDR). Signals being orders of magnitude apart makes it numerically difficult for deep learning approaches to deconvolve complex spectra. Therefore, scaling methods for HDR data in other deep learning domains are common practice, for example, Mel frequency cepstral coefficients (MFCC) in audio [38]. However, in 1D NMR, analogous embeddings of HDR signals are missing.

We introduce two pre-processing methods to make large parameter ranges better amenable to deep learning. The first pre-processing algorithm is called dynamic scaling. The method reduces the dynamic scale of the input spectrum and enhances the local contrast. The approach is inspired by techniques from computer vision for edge detection [39] and histogram equalization techniques [40]. This dynamic scaling is accomplished by applying a signal-dependent local contrast enhancement filter, cf. Algorithm 1. Therefore, the local maximum filter K_{max_b} and minimum filter K_{min_b} of different kernel widths b, where b is in the set of kernel widths $\mathbb{B}$, are applied to the real part of the spectrum S. To avoid large jumps due to discontinuous minimum and maximum filter mapping, the filtered spectra S_{min} and S_{max} are smoothed with a normalized Gaussian kernel $\mathbb{G}(\mu = 0, \sigma = b)$. Then the local minimum filtered spectrum S_{min_g} is subtracted from the original spectrum S and scaled by the local dynamic range, that is, the local maximum filtered minus the local minimum filtered spectrum
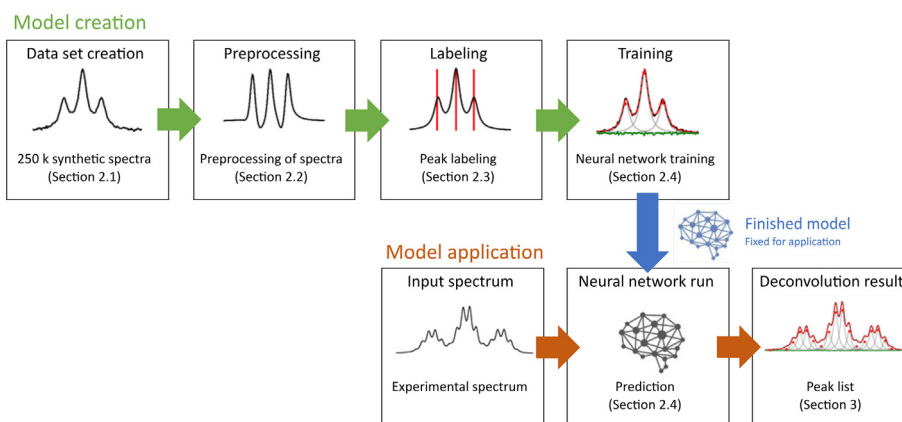

Fig. 2. Workflow diagram of model training and model application to experimental data. The prediction step includes pre-processing, inference, post-processing, and fitting.
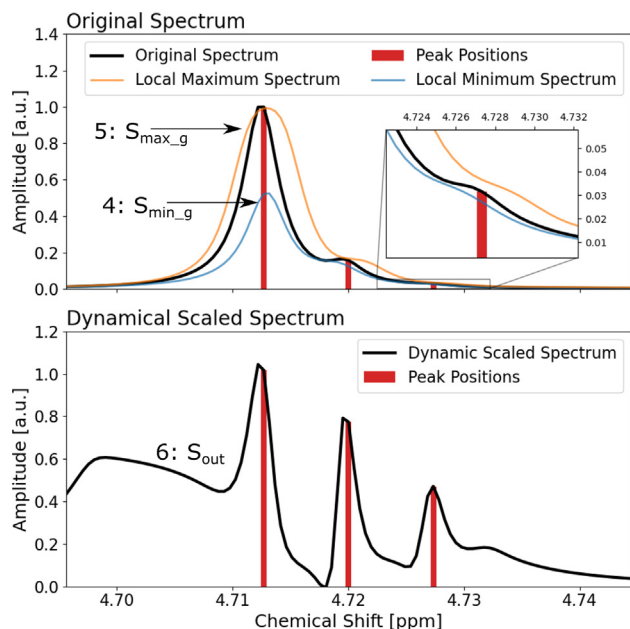
Fig. 3. Dynamic scaling for pre-processing the original spectrum (top, black) with three peaks (red). The procedure enhances the contrast by shifting the original spectrum with the local minimum spectrum (top, blue, cf. line number 4 Algorithm 1) and and by scaling with the local contrast, i.e., its local maximum (top, orange, cf. line number 5 Algorithm 1) minus its local minimum, making smaller peaks better detectable (bottom, black, cf. line number 6 Algorithm 1).

$S_{max\_g} - S_{min\_g}$. In this way, the dynamic range of the spectrum S_{out} is transformed approximately into the range $[0, 1]$, whereas minor local changes in amplitudes are enhanced; see Fig. 3.

Algorithm 1. Dynamic Scaling

1: **for** $b \in \mathbb{B}$ **do**	▷ iterate over set of kernel width
2: $S_{min} \leftarrow K_{min\_b}(S)$	▷ apply minimum filter
3: $S_{max} \leftarrow K_{max\_b}(S)$	▷ apply maximum filter
4: $S_{min\_g} \leftarrow S_{min} * G$	▷ smooth with Gaussian
5: $S_{max\_g} \leftarrow S_{max} * G$	▷ smooth with Gaussian
6: $S_{out} \leftarrow (S - S_{min\_g})/$	▷ shift and scale
$(S_{max\_g} - S_{min\_g})$	
7: **end for**	

Another variable with values in a wide range is the width of the peaks. On the one hand, there are peaks in 1D NMR spectra that have a width in the sub-Hz regime. On the other hand, exchangeables can have a width on the order of dozens of Hz, leading to spatial dependencies over dozens of Hz. Since we rely on the accuracy of deep learning in terms of position, we cannot use frequency resolution downsampling in the DNN. Having filters in the neural network that cover such a wide range adds more trainable parameters, that is, network weights, and hence entails a higher computational burden for training the network. Therefore, our method relies on shifting the spectrum. Instead of one spectrum being used as input to the network, several shifted versions of the same spectrum are fed into the network. This pre-processing-based solution increases the receptive field of the convolutional neural network (CNN) [41,42] part of the network. The receptive field is the part of the input spectrum to which a single output node is connected. Increasing the receptive field enables detecting broad signals.

2.3. Implicit regularization through labeling

Employing the peak parameters used to create the synthetic spectra is not a sensible option. Due to the inconclusive ways in which a spectrum can be reconstructed using different peak lists (cf. Fig. 1), the labeling would be ambiguous and inconsistent, making it difficult to train the neural network. Since synthetic spectra are utilized for training, an automatic labeling procedure is used for the data. Labeling the spectra in the case of machine learning methods corresponds to explicit regularization in classical fitting-based approaches. We use an automatic annotation for the position and width of the peaks. However, the linear parameters of the amplitude and combination between the Gaussian and Lorentzian lineshape are not learned but fitted afterwards. Therefore, we rely on a custom labeling approach.

Our labeling of the position of the peaks distinguishes between what should be detected and what not, or, in terms of performance, what is feasible to detect given the uncertainty in the experimental data. For this labeling technique, the width of lines of the original spectrum is shrunk by a factor that depends on the SNR in the original spectrum, cf. Fig. 4 (top and middle), and SI 5.4. The shrinking factor has a trade-off, cf. SI 5.4 for an illustration of the effect of different shrinking factors. On the one hand, not shrinking the linewidths leads to picking the local maxima as peak position labels, which is oversimplified. On the other hand, maximal shrinking results in restoring the original peak list, making the network potentially oversensitive toward minor distortions in the lineshapes. However, note that the recreation of the spectrum with the shrunk lines for the labeling is done without noise for accuracy reasons. The local maxima of the shrunk spectrum are then harnessed as labels for the original spectrum; see Fig. 4 (middle). The peak position labeling can be viewed as a local maximum picking in an altered, i.e., shrunk with respect to linewidths, spectrum. Similarly to the positions, we cannot use the widths from the synthetic spectrum generation since overlapping peaks will result in the same inconsistencies of the trained network. We also do not want to use an expensive fitting algorithm on the full spectrum to obtain reasonable estimates for the line widths. The present approach reduces the problem to a system of equations with $2n$ unknowns of amplitude and width and $2n$ equations at the n peak positions. Using an iteration approach to solve the problem reduces the equations even further to n block 2×2 systems that can, in special cases, be solved analytically, cf. SI 5.5 for technical details.

2.4. Translational-invariant convolutional and sequential deep learning

We use deep learning to find peak positions and widths. More precisely, we use one network for different but related tasks, also known as multitask learning. The network output determines for every point in the input spectrum whether the point represents a peak position or not. There is one no-peak class called baseline class and two peak classes. The peaks are divided into the two categories narrow peaks and broad peaks, i.e. exchangeables. At each peak position, the network further regresses the width of the corresponding peak.

We use a CNN for the extraction of local features and a type of recurrent neural network, a bidirectional long-short-term memory (LSTM), [43,44] to relate local characteristics to each other, cf. Fig. 5. 10 channels are fed into the network, that is, 2 dynamic scaling channels (kernel width in points $b \in \mathbb{B} = \{128, 2048\}$) and 8 shifting channels with a relative offset of 128 points, that is, approximately 25 Hz. The CNN starts with an inception layer [45] which is capable of extracting features of different bandwidths
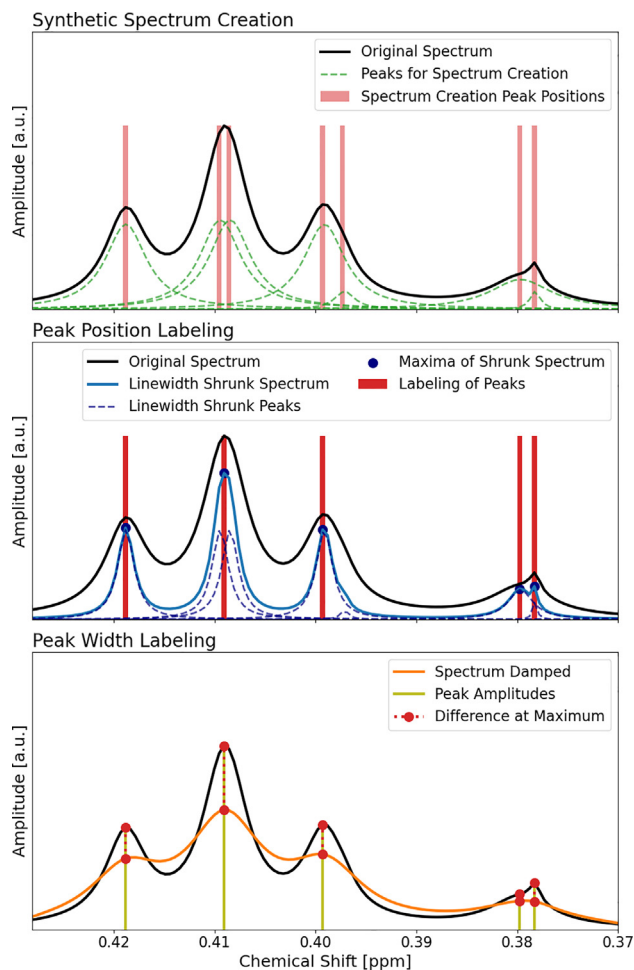
Fig. 4. Data Labeling Process: Synthetic Spectrum Creation (top panel): In the first step, peaks (top, green) are sampled. Their profiles are given in dashed green lines, while the position of the peaks is displayed in light red. All the lines together build the original spectrum (black). Position labeling process on synthetic spectra (center panel): The sampled peaks are shrunk relatively in width, by a SNR dependent factor (here 0.4) for the displayed peaks (dark blue dashed), and summed to a line shrunk spectrum (middle, light blue). In that line shrunk spectrum, the local maxima (dark blue dots) yield the positions of the frequency labels (red bars). Width labeling process (bottom panel): The original spectrum (black) is convolved with a Lorentzian peak to an exponential line broadened spectrum (orange). The two amplitudes (olive) of the original and the respective line broadened peak with their maxima (red dots) and the difference between the latter maxima (red dotted line) allow us to derive the width of the labeled peak through a system of equations.

Fig. 5. Data flow diagram for the neural network: This data flow diagram shows the neural network architecture and its corresponding inputs and outputs. First, the input spectrum (top panel) is pre-processed, leading to the multi-channel feature encoding (green panel) of the input spectrum. The spectral features are then fed into the translational-invariant neural network which yields three class outputs, i.e., top three rows for baseline, narrow peak and broad peak categories (black = true, white = false), and two regression outputs, i.e., bottom two rows for narrow peak width and broad peak width (greyscale for width within class bounds). The network output is post-processed and the parameters are fitted using the post-processed neural network output as an initial guess. This results in a peak list through which we can reconstruct the input spectrum (bottom panel).

simultaneously, cf. the detailed architecture and hyperparameter settings in SI 5.6. The inception layer connects over time-distributed dense layers to the bidirectional LSTM. The bidirectional LSTM is responsible for relating the local features extracted by the previous convolutional and dense layers and setting them into a more global perspective. Thus, the network should be able to relate and use information from peaks dozens of Hz apart, e.g. from multiplets with high coupling constants. The output of the LSTM connects over another sequence of time-distributed dense layers to the 5 output channels. These 5 channels produce for each point on the frequency axis three classes of baseline, narrow peaks, and broad peaks, and their respective peak widths.

The whole network is designed in such a way that it is translational-invariant [46]. This means that we are not restricted to a particular spectral width, and the network is equivariant towards the position of a peak in the input spectrum. Therefore, the network applies the same mapping and thus gives the same output for a specific region independent of the chemical shift of the former region, e.g. whether the region is at 1 ppm or at
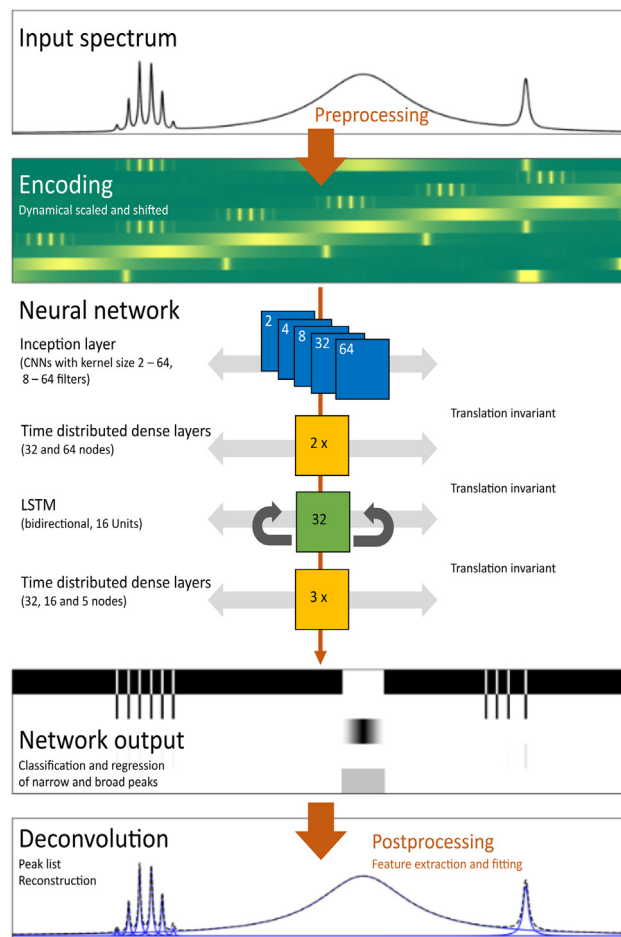
5 ppm. A data set of 250'000 synthetic spectra with 8192 points each was created and used for training. This allows the fitting of broad features, such as broad peaks, into a training sample. Network training is performed using the pre-processed spectrum as input and our labeling as target. The width of the binary position labels in discrete points is 3 for the narrow peak class and is proportional to the width for the broad peak class. We use a binary cross-entropy loss for the classification tasks (output channels 1–3) and a mean squared error (MSE) loss for the regression (output channels 4–5). Classification and regression losses are equally weighted. Logarithmic scaling for the broad peak class is applied to compactly represent the wide range of width values. All regression parameters are normalized to the same order of magnitude and thus stabilize the training process numerically. Note that for nonpeak positions, the regression loss is masked to avoid contamination of the error. The purpose of deep learning is to obtain the correct number of peaks, that is, the correct order of the model, and a promising valley, i.e. proximity of a local minimum, for further convex peak parameter fitting.

To apply a trained network to the experimental spectra, the experimental spectra are phase and baseline corrected, line broad-

ened and resampled, see Section 2.1. Afterward, the experimental spectra are pre-processed with the dynamic scaling filter and the shifting method in the same way as the training data, cf. Section 2.2. These pre-processed versions of the spectrum are then fed into the neural network. The output of the neural network is post-processed with the help of the baseline class and non-maximum suppression (NMS) [47] in order to obtain sparse peak positions. The neural network predictions are then used for a subsequent iterative classical fitting procedure, where the number of peaks is fixed. The individual peak parameters are sequentially fitted for each peak with a gradient-free optimization algorithm. The quality of the predictions of the neural network is crucial as it serves as an initial guess for the fitting.

3. Results and discussion

In this section, we demonstrate the deconvolution performance of our method on synthetic and experimental spectra. Furthermore, we discuss the methodological advantages of our approach compared to other techniques. The deconvolution result of the same spectrum may vary from expert to expert, especially if they cannot rely on additional structural information. Thus, we demonstrate our results in awareness of a remaining ambiguity regarding an optimal solution. The main application of our method is 1D high-field ^{1}H spectra. However, the algorithm is generally not restricted to this use case, hence applications in other situations may benefit from retraining or fine-tuning the network.

3.1. Statistical deconvolution results on synthetic data

The deconvolution performance statistics on synthetic data, cf. details on synthetic testset supplementary information (SI) 5.2,

are presented in Table 1 and Fig. 6A. The statistics are divided into three subcategories: Picking Accuracy Score $M_1 \in [0, 1]$, Overpicking/Sparsity Score $M_2 \in [0, 1]$ and Reconstruction Score $M_3 \in [0, 1]$. These three criteria quantify what we consider to be a good deconvolution result. Together, these three scores yield a Total Score $M_{tot} = M_1 \cdot M_2 \cdot M_3 \in [0, 1]$ for each test region; cf. SI 5.8 for details on the metrics. The results on synthetic data are excellent with high values in all three individual scores M_1, M_2, and M_3, and also in the total score M_{tot}. Since the synthetic test spectra possess minor phase and baseline distortions, line shape distortions, that is, Voigt lines, and solvent peaks (cf. SI 5.2), we conclude that we have achieved a certain robustness toward small distortions. Furthermore, the over-picking score M_2 is worth highlighting, with a perfect score of $M_2 = 1$ for all synthetic test regions. This means that the network never predicts more peaks than peaks that were annotated, and hence yields a sparse peak list. Distortions and crowded regions, such as in the synthetic testset, generally make deconvolution more arduous and can lead to performance degradation; cf. SI 5.10, 5.11, and 5.12. However, overall, we observe high scores in all categories despite distortions and overlapping signals.

3.2. Statistical deconvolution results on experimental data

The deconvolution performance statistics on experimental data, cf. for details of the experimental test set in SI 5.3, are presented in Table 2 and Fig. 6B. The median of the overall score M_{tot} on the experimental data is only 8% lower than on the synthetic data and therefore high at 82%. Due to the distribution shift between the synthetic and experimental data, a minor performance degradation on the experimental data is expected. However, in the peak picking M_1 and reconstruction score M_3, the median value on

Table 1
Statistical key figures for the deconvolution results on synthetic data with respect to the metrics.

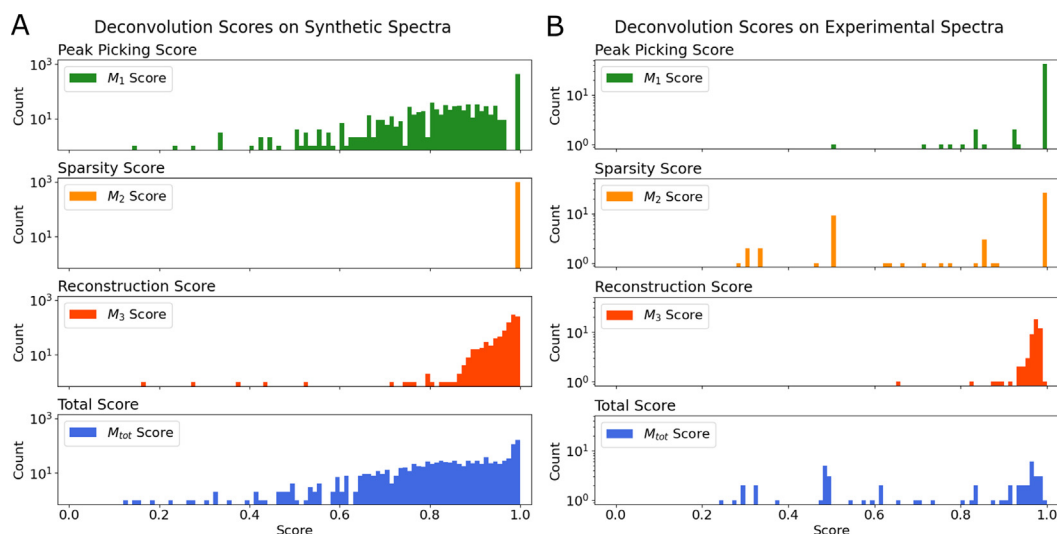| Metrics	Deconvolution on Synthetic Data			
	Lower Quartile Q_1	Median $\bar{\mu}$	Upper Quartile Q_3	Score > 0.9
Picking Accuracy Score M_1	0.82	0.92	1.00	54.8%
Overpicking/Sparsity Score M_2	1.00	1.00	1.00	100%
Reconstruction Score M_3	0.96	0.98	0.99	95.4%
Total Score M_{tot}	0.80	0.90	0.98	49.5%

Fig. 6. (A) Deconvolution scores on 1000 synthetic regions: The deconvolution performance statistics on synthetic test regions are represented in the first 4-panel column. (B) Deconvolution scores on 53 experimental regions: In the second column of 4 panels, the performance statistics on experimental test regions are displayed.

Table 2

Statistical key figures for the deconvolution results on experimental data with respect to the metrics.

Metrics	Deconvolution on Experimental Data			
	Lower Quartile Q_1	Median $\bar{\mu}$	Upper Quartile Q_3	Score > 0.9
Picking Accuracy Score M_1	1.00	1.00	1.00	84.9%
Overpicking/Sparsity Score M_2	0.50	0.88	1.00	49.1%
Reconstruction Score M_3	0.96	0.97	0.98	90.6%
Total Score M_{tot}	0.49	0.82	0.96	41.5%

experimental regions is even better than on synthetic data. This can be explained by the higher crowdedness and more substantial distortions, i.e., deconvolution difficulty, of the synthetic spectra. Sparser and less distorted synthetic data sets lead to superior results for M_1 and M_3 compared to the deconvolution of the experimental test set, cf. SI 5.12. On the other hand, the median of the sparsity score M_2 on experimental data is lower than on synthetic data, cf. Tables 1 and 2. A factor for this discrepancy is the smaller peaks, e.g. satellites, that were ignored by the experts but not by the algorithm (cf. SI 5.9 for how to discard low SNR peaks). We conclude that the network trained on synthetic data can be well transferred to experimental data, since the differences in the scores are minor and well-explainable.

3.3. Ambiguity and limitations of deconvolution: A comparison with experts

Crowded regions are challenging to deconvolve, since peaks are easier to distinguish if they do not overlap. Therefore, the stronger a region is crowded, the more diverse are the expert deconvolution results. For three high-field ^{1}H regions with different amounts of overlapping peaks, we compare the results of our method, cf. Fig. 7 and cf. Table 3, with the manual results of 10 experts, cf. Table 3. The experts have tried to detect the lines and adjusted the peak parameters (frequency, amplitude, width, and line shape) to minimize the residual upon visual inspection. Note that these experts did not have additional structural information. In terms

of the number of picked peaks, we gave the minimum, maximum, and median numbers of the expert group. For the simplest region A, the experts agree on the number of peaks, whereas for the most challenging region C the expert annotations vary between 14 and 24 peaks. The result of our method in terms of the number of peaks matches the median of the experts for two of three regions. In terms of residual, we calculated the average and standard deviation (Std) of the expert group's MAE. The largest MAE and Std of the MAE are observed for region C. The results of the expert deconvolutions exhibit significant discrepancies for crowded regions regarding peaks picked and reconstruction error.

The results of our method for these three regions lie within the expert group's range in picked peaks and yield a lower residual than most experts, cf. Table 3. The mean absolute error MAE of our method lies at least one Std lower than the average of the experts for all three regions. Thus, we observe good transferability to sparse and crowded regions in experimental data by the example of three different regions and thus, can cover a wide range of deconvolution problems. Note that the result of our method in the most crowded region, cf. Fig. 7 C, is especially interesting, since there might be significantly fewer local maxima than lines that generate the region. The peak list of this crowded region is sparse, but the solution is not oversimplified, e.g., two doublets between $0.33 - 0.35$ ppm are found, despite observing only two local maxima. Altogether, these results suggest that our method can keep up with and, in terms of MAE, even substantially outperform the manual deconvolution results of experts.
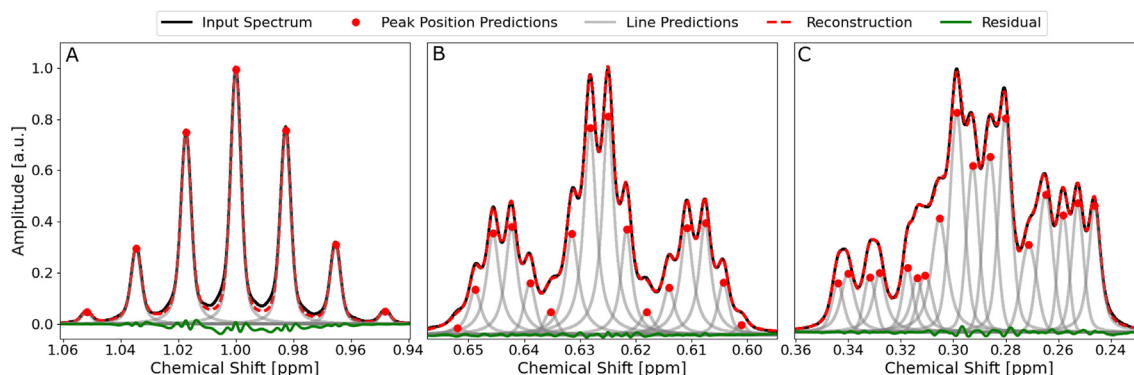

Fig. 7. This three deconvolution results display the performance of the method from sparse (A) to strongly crowded (C) regions of experimental ^{1}H high-field spectra. The black line indicates the input spectrum to the algorithm. The red dots indicate the predicted peak positions (frequencies) and amplitudes, whereas the corresponding lines are displayed in grey. The dashed red line depicts the reconstructed spectrum and the green line the residual error, i.e., the difference between the input and reconstructed spectrum.

Table 3

We compared the deconvolution results of our method with the expert results of three regions, A, B, and C, in Fig. 7.

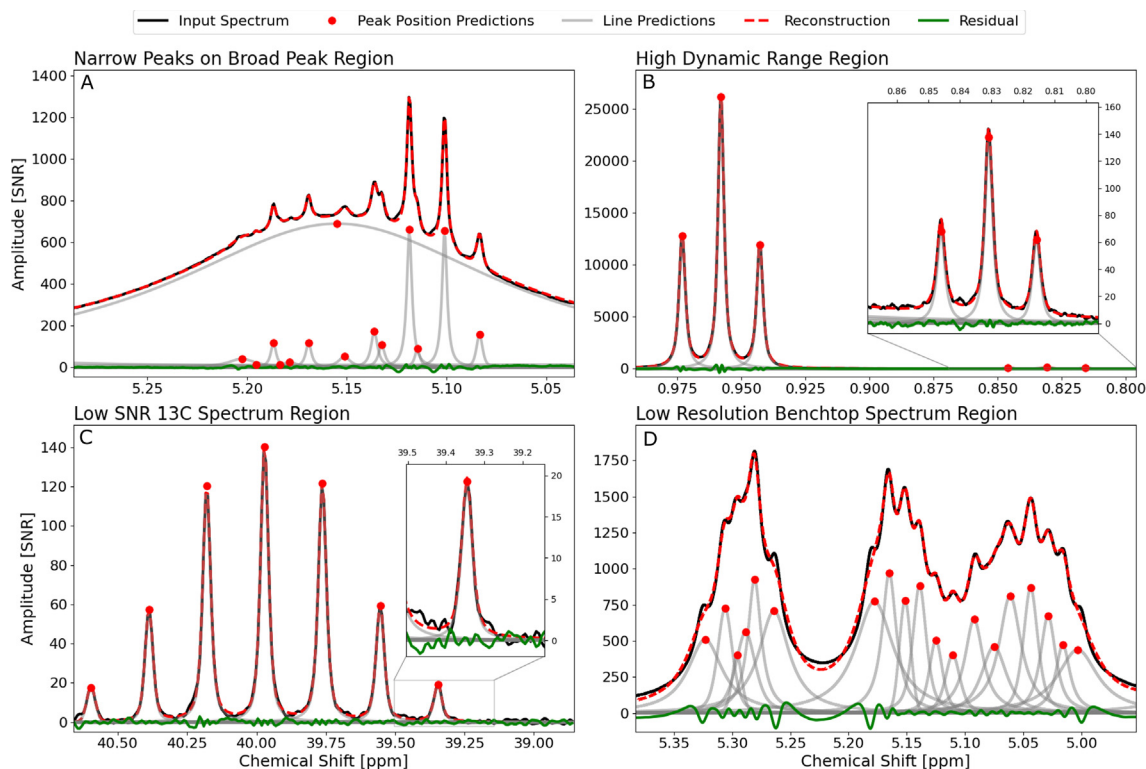| Regions	Experts					Our Approach	
	Min # Peaks	Max # Peaks	Median # Peaks	Average MAE	Std MAE	# Peaks	MAE
A	7	7	7	0.0067	0.0028	7	0.0034
B	16	18	18	0.0067	0.0036	16	0.0029
C	14	24	17	0.0141	0.0116	17	0.0025

Fig. 8. Deconvolution results of four experimental regions: (A) Narrow peaks on broad peaks: This deconvolution results from an experimental spectrum with a region where broad lines, i.e., exchangeables, overlap with narrow peaks. (B) High dynamic range region: This deconvolution result of an experimental ^{1}H spectrum region illustrates how the algorithm copes with a high-dynamic range. A triplet (left) and one of its satellites (zoom, right) are shown. (C) Low SNR ^{13}C region: The SNR 10 peak at 39.35 ppm is reconstructed accurately. (D) Benchtop region: We get a sparse and accurate deconvolution despite low resolution and shoulder peaks in a spectrum acquired on a low-field benchtop NMR system.

3.4. Detailed example regions

In this section, applications of our method are illustrated by four different regions, cf. Fig. 8. In the example of the first region with broad and narrow peaks that overlap, we observe excellent results of our method in terms of low residual and sparsity of the peak list, cf. high-field ^{1}H region in Fig. 8A. By separating the prediction of broad peaks from the prediction of narrow peaks in the neural network, the algorithm can disentangle overlapping lines of broad and narrow peaks that are orders of magnitude apart in terms of their width.

Features orders of magnitude apart pose challenges to neural networks since different scales of inputs cause different parameter updates and optimizer steps towards the minimum of the loss function. Our method enables good deconvolution results for spectra of a very high dynamic range. Hence, the algorithm does not miss low SNR peaks. The second showcase region illustrates this ability, see high-field ^{1}H region Fig. 8B. Although the magnitudes of the peaks are orders apart, the reconstruction is nearly perfect and the residual small, also compared to the peak intensities.

The third example shows a ^{13}C spectrum region. ^{13}C nuclei have a lower natural abundance than ^{1}H nuclei, usually resulting in lower sensitivity. Although our method focuses on ^{1}H high-field spectra, we can detect and accurately fit low SNR peaks of ^{13}C spectra, cf. ^{13}C region in Fig. 8C.

The last showcase displays a spectrum acquired on a low-field (80 MHz) Benchtop NMR system, cf. ^{1}H Benchtop region in Fig. 8D. Benchtop spectra are challenging for deconvolution due to their intrinsic lower resolution, i.e., the large ratio of line widths to coupling constants, which often leads to overlapping peaks.

Furthermore, the ratio of the coupling constant to the difference in frequency is much smaller in Benchtop spectra resulting in highly asymmetric peaks through strong coupling.

3.5. Discussion on training with synthetic vs. experimental data

1D NMR data can have strongly divergent properties: Single peaks can have various shapes, widths, and SNRs. Furthermore, spectral regions can be sparse or crowded, and they can have a low or a high dynamic range. This often leads to an underrepresentation of specific spectral characteristics in experimental NMR data sets. Hence, we think that relying on synthetic data, and therefore being able to shape the data distribution, is a significant advantage. Using experimental data sets, one must necessarily deal with imbalanced data, which makes it difficult to solve rare but important region types [17,48].

Spectra annotated by experts suffer from label noise. Label noise refers to variations in quality and the lack of consistency in the expert annotation of experimental spectra, cf. Section 3.3. Hence, label noise can hamper the training convergence, since the algorithm does not have a unique reference what it should learn, e.g. two identical regions are labeled differently. This problem becomes especially obvious in regions with strongly overlapping peaks, resulting in a respective performance drop.

Compared to other synthetic data-based machine learning methods for peak picking and deconvolution in NMR, which randomly sampled single peaks [31], we model signals through multiplet structures. This bears the benefit that peak parameters are shared within the same multiplet, e.g., width, coupling constants, and intensity ratios, and can be learned and exploited by the deep learning model.

3.6. Discussion on extracting HDR and broad features

Our pre-processing, namely the dynamic scaling and shifting approach, essentially supports the neural network to get the data needed to detect peaks in these challenging situations, see SI 5.13 for a quantitative analysis on the benefits of the dynamic scaling. Due to the augmented wide field of view and the bidirectional LSTM, the network can capture more far-reaching dependencies, which are needed e.g., to determine the width of broad peaks or to extract shared information in multiplets with high coupling constants. However, other deep learning algorithms for peak picking or deconvolution in NMR rely solely on local CNN-based feature extraction and a single peak class [31,49]. A narrow field of view combined with a single peak class cannot handle broad features and the overlap of broad and narrow peaks, for example, see Fig. 8A.

3.7. Discussion on pattern recognition for Ill-posed deconvolution problem

Classical fitting-based 1D deconvolution approaches already exist and are frequently used; cf. Chapter 1. It is shown that it is unnecessary to solve linear parts of the optimization problem, for instance, for the amplitude and Lorentzian-to-Gaussian fraction parameter of the pseudo-Voigt lines, with deep learning. These cases can be solved with less computational cost and more accurately with convex solvers [50]. However, particularly difficult parts of the deconvolution problem greatly benefit from deep learning. Therefore, our approach relies on deep learning for critical components of the inverse problem of deconvolution, namely the integer optimization problem of finding the number of peaks and the non-convex program for determining their corresponding position and width. The reasons for this superiority of deep learning approaches for these tasks are manifold. On the one hand, to get the proper model order with classical optimization algorithms, one has to trade off the fitting error and the number of peaks explicitly. This often does not work robustly in complex scenarios [3]. On the other hand, circumventing an explicit regularization criterion with labeling, and thus reducing it to a pattern recognition problem, has several advantages. Through labeling, one can essentially teach the network what a peak is and what not. In particular, one can make use of regular patterns in NMR. The deconvolution solution becomes tunable by adapting the line shrinking parameter in the labeling; cf. SI 5.4. In contrast, in classical fitting-based deconvolution methods, this would often mean optimizing explicit regularization parameters whose implications on the deconvolution result are difficult to assess. For example, we compare a classical with a deep learning-based deconvolution algorithm, cf. Fig. 9. The classical algorithm is oriented on Mestre-GSD [15], but not an exact implementation of it. The deep learning algorithm is our approach. We manually tuned the parameters of the classical algorithm to yield good results in two specific regions. However, this also demonstrates a downside of many classical algorithms, i.e., the need for region- or spectrum-specific tuning of parameters. Furthermore, we observed superior results of our approach, especially on shoulder and broad peaks, which is demonstrated in Fig. 9. In the upper panel of Fig. 9, a rather crowded region is shown. Our method finds shoulder peaks more reliably, e.g., zoom at 5.12 ppm and 4.89 ppm, and does not over-pick in contrast to the classical picker, e.g., zoom at around 5.05 ppm for a false picked peak. In the second panel of Fig. 9, a broad peak is exhibited. We observe that the classical algorithm has problems with the broad peak and picks several peaks close to the maximum of the broad peak. In contrast our method yields a sparse single peak solution for this region. Additionally, having a reasonable estimate of the nonlinear parameters and model order
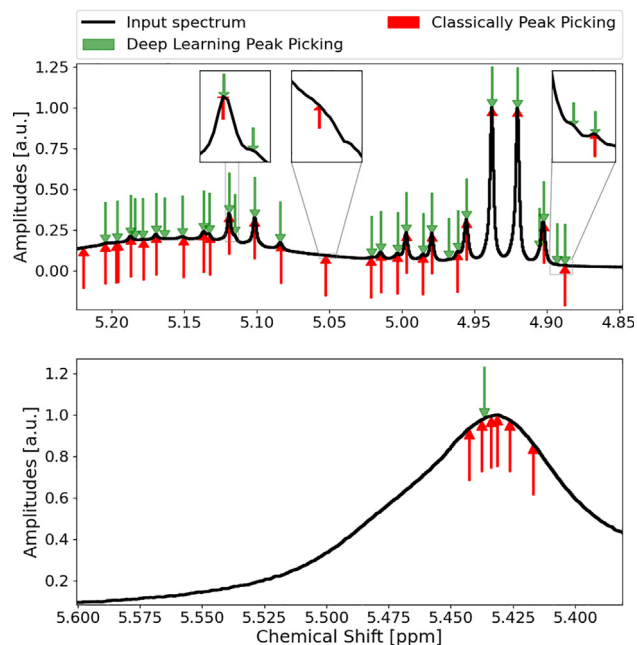

Fig. 9. Comparison between a classical and our deep learning-based deconvolution algorithm. The spectra in the two displayed regions (top panel and bottom panel) are drawn in black, whereas the picked peaks are indicated with arrows pointing down in green for our approach and arrows pointing up in red for the compared classical algorithm.

through the prediction of the neural network reduces the rest of the optimization problem often to a locally convex fine-tuning, which can be solved efficiently. The deconvolution of a reasonably crowded spectrum, i.e. more than 500 peaks, is completed in a few seconds.

For efficient training and good deconvolution performance, consistent labeling is essential. In contrast to labeling methods based on line refitting of peaks, cf. Li et al. [31], our labeling approach is fast and, therefore, can be computed on the fly. Compared to labeling based on local maxima, the present methods can detect multiple peaks even if there is just a single local maximum within the region.

4. Conclusion

We introduce a robust, expert-level quality deep learning-based deconvolution algorithm for 1D experimental NMR spectra. Our custom network, labeling, and pre-processing enable excellent performance in challenging scenarios such as crowded, high dynamic range and shoulder peak regions, as well as in cases of overlap of narrow with broad peaks. On experimental high-field ^{1}H spectra, our algorithm yields low residual errors between the reconstructed spectrum and the original spectrum and sparse peak lists. The number of peaks proposed by our algorithm and the residual error lie within the uncertainty range of expert annotations. Furthermore, the algorithm copes well with Gaussian line-shape distortions, i.e. Voigt lines, and small baseline and phase distortions, reflected by good test results with respect to the number of peaks picked and residual on distorted synthetic spectra. Thus, the presented algorithm is at least comparable, if not superior, to human experts and robust toward small distortions. Although developed for high-field proton spectroscopy, we find that our method can process different nuclei, base frequencies, resolutions, and sensitivities. Altogether, by proving its usefulness in challenging situa-

tions, our approach is an attractive 1D NMR deconvolution algorithm.

Further testing robustness against substantial distortions, quantifying the uncertainty of predictions, and interpreting the neural network model could be interesting future research directions. Moreover, we are planning to extend our research towards the extraction of multiplet information, which assists structure elucidation and verification.

Data Availability

Experimental data will be available upon request.

Method Availability

The method is available within the TopSpin TS 4.2 release of Bruker Biospin with the command "mldcon".

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: All authors except Jan Dirk Wegner, Roland Sigel and Helmut Grabner report financial support was provided by Innosuisse Swiss Innovation Agency. All authors from Bruker Switzerland AG report a relationship with Bruker BioSpin AG that includes: employment.

Acknowledgment

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jmr.2022.107357.

References

[1] D. Chen, Z. Wang, D. Guo, V. Orekhov, X. Qu, Review and Prospect: Deep Learning in Nuclear Magnetic Resonance Spectroscopy, Chem. – A Eur. J. 26 (2020) 10391–10401.
[2] A.N. Tikhonov, Ill-Posed Problems in Natural Sciences: Proceedings of the International Conference Held in Moscow, August 19–25, 1991.
[3] P. Stoica, Y. Selen, Model-order selection: a review of information criterion rules, IEEE Signal Process. Mag. 21 (2004) 36–47.
[4] A. Liavas, P. Regalia, On the behavior of information theoretic criteria for model order selection, IEEE Trans. Signal Process. 49 (2001) 1689–1695.
[5] R. Koradi, M. Billeter, M. Engeli, P. Güntert, K. Wüthrich, Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY, J. Magn. Reson. 135 (1998) 288–297.
[6] B.U. Meier, Z.L. Madi, R.R. Ernst, Computer analysis of nuclear spin systems based on local symmetry in 2D spectra, J. Magn. Reson. 74 (1987) 565–573.
[7] Z. Liu, A. Abbas, B.-Y. Jing, X. Gao, WaVPeak: picking NMR peaks through wavelet-based smoothing and volume-based filtering, Bioinformatics 28 (2012) 914–920.
[8] S. Tikole, V. Jaravine, V. Rogov, V. Dötsch, P. Güntert, Peak picking NMR spectral data using non-negative matrix factorization, BMC Bioinformat. 15 (2014) 46–53.
[9] B. Alipanahi, X. Gao, E. Karakoc, L. Donaldson, M. Li, PICKY: a novel SVD-based NMR spectra peak picking method, Bioinformatic 25 (2009) i268–275.
[10] Y. Cheng, X. Gao, F. Liang, Bayesian Peak Picking for NMR Spectra, Genomics, Proteom. Bioinformat. 12 (2014) 39–47.
[11] K. Krishnamurthy, CRAFT (complete reduction to amplitude frequency table) - robust and time-efficient Bayesian approach for quantitative mixture analysis by NMR: CRAFT - complete reduction to amplitude frequency table, Magn. Reson. Chem. 51 (2013) 821–829.
[12] M. Anjum, P.A. Dmochowski, P.D. Teal, A subband Steiglitz-McBride algorithm for automatic analysis of FID data, Magn. Reson. Chem. 56 (2018) 740–747.
[13] N. Sandgren, Y. Selén, P. Stoica, J. Li, Parametric methods for frequency-selective MR spectroscopy-a review, J. Magn. Reson. 168 (2004) 259–272.
[14] D. Belkić, K. Belkić, Fast Padé transform for optimal quantification of time signals from magnetic resonance spectroscopy, Int. J. Quantum Chem. 105 (2005) 493–510.
[15] C. Cobas, F. Seoane, S. Sykora, Global Spectral Deconvolution (GSD) of 1D-NMR spectra, Poster, Poster, SMASH Conference, 2008.
[16] C.A. Reilly, B.R. Kowalski, Nuclear magnetic resonance spectral interpretation by pattern recognition, J. Phys. Chem. 75 (1971) 1402–1411.
[17] C. Cobas, NMR signal processing, prediction, and structure verification with machine learning techniques, Magn. Reson. Chem. 58 (2020) 512–519.
[18] J.A. Fine, A.A. Rajasekar, K.P. Jethava, G. Chopra, Spectral deep learning for prediction and prospective validation of functional groups, Chem. Sci. 11 (2020) 4618–4630.
[19] S.G. Worswick, J.A. Spencer, G. Jeschke, I. Kuprov, Deep neural network processing of DEER data, Sci. Adv. 4 (2018) eaat5218.
[20] F.M. Paruzzo, A. Hofstetter, F. Musil, S. De, M. Ceriotti, L. Emsley, Chemical shifts in molecular solids by machine learning, Nat. Commun. 9 (2018) 4501.
[21] B. Han, Y. Liu, S.W. Ginzinger, D.S. Wishart, SHIFTX2: significantly improved protein chemical shift prediction, J. Biomol. NMR 50 (2011) 43–57.
[22] Y. Shen, A. Bax, SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network, J. Biomol. NMR 48 (2010) 13–22.
[23] S. Liu, J. Li, K.C. Bennett, B. Ganoe, T. Stauch, M. Head-Gordon, A. Hexemer, D. Ushizima, T. Head-Gordon, Multiresolution 3D-DenseNet for chemical shift prediction in NMR crystallography, J. Phys. Chem. Lett. 10 (2019) 4558–4565.
[24] D. Li, R. Brüschweiler, PPM_One: a static protein structure based chemical shift predictor, J. Biomol. NMR 62 (2015) 403–409.
[25] M.A. Zambrello, D.L. Craft, J.C. Hoch, D. Rovnyak, A.D. Schuyler, The influence of the probability density function on spectral quality in nonuniformly sampled multidimensional NMR, J. Magn. Reson. 311 (2020) 106671.
[26] D.F. Hansen, Using deep neural networks to reconstruct non-uniformly sampled NMR spectra, J. Biomol. NMR 73 (2019) 577–585.
[27] X. Qu, Y. Huang, H. Lu, T. Qiu, D. Guo, T. Agback, V. Orekhov, Z. Chen, Accelerated nuclear magnetic resonance spectroscopy with deep learning, Angew. Chem. Int. Ed. 59 (2020) 10297–10300.
[28] H.H. Lee, H. Kim, Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain, Magn. Reson. Med. 82 (2019) 33–48.
[29] P. Klukowski, R. Riek, P. Güntert, Rapid protein assignments and structures from raw NMR spectra with the deep learning technique ARTINA, Nat. Commun. 13 (2022) 6151.
[30] Y. Huang, J. Zhao, Z. Wang, V. Orekhov, D. Guo, X. Qu, Exponential signal reconstruction with deep hankel matrix factorization, IEEE Trans. Neural Networks Learn. Syst. (2021) 1–13.
[31] D.-W. Li, A.L. Hansen, C. Yuan, L. Bruschweiler-Li, R. Brüschweiler, DEEP picker is a deep neural network for accurate deconvolution of complex two-dimensional NMR spectra, Nat. Commun. 12 (2021) 5229.
[32] D.-W. Li, A.L. Hansen, L. Bruschweiler-Li, C. Yuan, R. Brüschweiler, Fundamental and practical aspects of machine learning for the peak picking of biomolecular NMR spectra, J. Biomol. NMR 76 (2022) 49–57.
[33] J. Quinonero-Candela, M. Sugiyama, A. Schwaighofer, N. Lawrence, Dataset Shift in Machine Learning, Neural Information Processing series, MIT Press, 2008.
[34] S.J. Pan, Q. Yang, A Survey on Transfer Learning, IEEE Trans. Knowl. Data Eng. 22 (2010) 1345–1359.
[35] J. Olivier, S. Kilani, R. Poirier, Determination in low-energy electron loss spectroscopy of the Gaussian and Lorentzian content of experimental line shapes, Appl. Surface Sci. 8 (1981) 353–358.
[36] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (2019) 60.
[37] S. Bruderer, F. Paruzzo, C. Bolliger, Deep learning-based phase and baseline correction of 1D 1H NMR Spectra, Public Bruker White Paper, URL: https://www.bruker.com/en/products-and-solutions/mr/nmr-software/topspin.html, 2021.
[38] S.K. Pandey, H.S. Shekhawat, S.R.M. Prasanna, in: Proceedings of the 29th International Conference Radioelektronika, 29th International Conference Radioelektronika, 2019, pp 1–6.
[39] J. Canny, A Computational Approach to Edge Detection, IEEE Trans. Pattern Anal. Mach. Intell., 1986, pp. 679–698.
[40] S. Pizer, R. Johnston, J. Ericksen, B. Yankaskas, K. Muller, In proceedings of the first conference on visualization in biomedical, Computing (1990) 337–345.
[41] A. Krizhevsky, I. Sutskever, G.E. Hinton, in: F. Pereira, C.J. Burges, L. Bottou, K.Q. Weinberger, (Eds.), Advances in Neural Information Processing Systems, vol. 25, Curran Associates Inc, 2012.
[42] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[43] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, Neural Comput. 9 (1997) 1735–1780.
[44] M. Schuster, K. Paliwal, Bidirectional recurrent neural networks, IEEE Trans. Signal Process. 45 (1997) 2673–2681.
[45] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp 1–9.
[46] R. Zhang, in: Proceedings of the 36th International Conference on Machine Learning, International Conference on Machine Learning, 2019, pp 7324–7334.

[47] A. Neubeck, L. Van Gool, in: Proceedings of the 18th International Conference on Pattern Recognitio, International Conference on Pattern Recognition, vol. 3, 2006, pp 850–855.

[48] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, G. Bing, Learning from class-imbalanced data: Review of methods and applications, Expert Syst. Appl. 73 (2017) 220–239.

[49] P. Klukowski, M. Augoff, M. Zięba, M. Drwal, A. Gonczarek, M.J. Walczak, NMRNet: a deep learning approach to automated peak picking of protein NMR spectra, Bioinformatics 34 (2018) 2590–2597.

[50] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, 2004.