

PEAK DETECTION AND BASELINE CORRECTION USING A CONVOLUTIONAL NEURAL NETWORK

Mikkel N. Schmidt Tommy S. Alstrøm Marcus Svendstorp Jan Larsen

Department of Applied Mathematics and Computer Science, Technical University of Denmark

ABSTRACT

Peak detection and localization in a noisy signal with an unknown baseline is a fundamental task in signal processing applications such as spectroscopy. A current trend in signal processing is to reformulate traditional processing pipelines as (deep) neural networks that can be trained end-to-end. A trainable algorithm for baseline removal and peak localization can serve as an important module in such a processing pipeline. In practical applications, one of the most successful approaches to joint baseline suppression and peak localization is based on the continuous wavelet transform: We reformulate this as a convolutional neural network (CNN) followed by a non-linear readout layer. On a synthetic benchmark we demonstrate that with sufficient training data, the CNN approach consistently outperforms the optimized continuous wavelet method by means of adapting to the spectral peak shape, noise level, and characteristics of the baseline. The CNN approach to peak localization shows great promise, as it can more efficiently leverage data to outperform the current state of the art, and can readily be extended and incorporated as a module in a larger neural network architecture.

Index Terms— Peak detection, Peak localization, Wavelet, Spectroscopy, Convolution Neural Network

1. INTRODUCTION

Peak detection is a vital component in a wide range of applications, and serves as a component in applications such as molecule identification in spectroscopy [1, 2], chromatography [3], beat and onset detection in audio [4], and event detection in social media data [5]. In spectroscopy, the characteristic peak locations comprise a fingerprint of a measured analyte [1]. In practice, a measured spectrum also contains an unknown smooth baseline component, and to reliably detect the peak locations in a measured spectrum, baseline correction (also called background removal) is often carried out.

A traditional procedure for baseline removal and peak localization typically consists of three steps [2]; 1) smoothing,

2) baseline correction and 3) peak picking. Numerous algorithms for these steps have been proposed in the literature: Smoothing can e.g. be carried out using splines, Savitzky-Golay filters [6], partial least squares [7], or undecimated wavelet transform [8]. Baseline estimation can also be carried out in multiple ways: The signal can be fitted using a sufficiently smooth model, such as a polynomial of low degree or a neural network with smooth basis functions [9], using a non-symmetric error function, so it tends to fit the baseline rather than the signal peaks. Alternatively, nonparametric methods such as monotone minimum, local regression [10], or moving average minimum/quantile filtering have been used to estimate the baseline. Finally, the estimated baseline is subtracted from the signal, and the peaks are located by a suitable peak picking procedure. Algorithms that jointly estimate the baseline and peak locations have also received attention recently [11, 3, 12], and have been shown to avoid some of the problems associated with the traditional two-step procedure.

As an alternative to the outlined estimation procedures, peak localization methods based on continuous wavelet transforms can automatically suppress the baseline [13]. In a comprehensive survey and experimental comparison of peak detection methods for mass spectrometry, Yang et al. conclude that “the continuous wavelet based algorithm provides the best average performance” [2].

In a typical signal processing pipeline, where several algorithms are applied to a signal in succession, errors can accumulate and it can be unclear how to tune the different algorithms to perform best in combination. A current trend in signal processing is to reformulate existing processing pipelines as neural network architectures, such that the complete pipeline can be trained end-to-end in a supervised manner. In this paper we demonstrate how the continuous wavelet based peak localization procedure with baseline suppression can be designed as a trainable neural network, allowing it to serve as a module in a larger processing network.

In section 2 we delimit the problem and review the continuous wavelet approach to peak localization. Next, we reformulate this approach as a trainable neural network, and introduce two benchmark methods. In section 3 we evaluate the proposed method in a large cross validated synthetic data study, and show that the proposed method consistently outperforms the continuous wavelet based approach, which we

This research was funded by the IDUN Center of Excellence supported by the Danish National Research Foundation (DNRF122) and the Velux Foundations (Grant No. 9301). We also thank NVIDIA corporation for donating GPU.

consider the current state of the art for an unsupervised algorithm. Finally, we discuss the results and possible extensions, and conclude in section 4.

2. METHOD

In this paper, we address the situation where a measured spectrum, $s(f)$, consists of the summation of an unknown smooth baseline signal, $b(f)$, and a single spectral peak with fixed but unknown lineshape, $v(f - f_0)$, situated at an unknown location, f_0 , where f is the spectral/frequency variable. The spectrum is contaminated by i.i.d. Gaussian noise, $e(f)$, scaled by a factor s to yield a specific peak signal-to-noise ratio,

$$s(f) = b(f) + v(f - f_0) + s \cdot e(f). \quad (1)$$

Given such a signal, the goal is to estimate the location of the peak. We assume a training set with multiple of such signals, each with different baselines, peak locations, and noise, is available to train the algorithm, which is evaluated by the mean absolute error (MAE) on a separate test set.

2.1. Continuous wavelet peak localization

In continuous wavelet peak localization, the measured spectrum is convolved by a suitable wavelet function, and the peak location is estimated as the frequency at which the convolution has maximum magnitude. When the lineshape of the peak can vary, a basis with multiple wavelets is beneficial, but in the fixed lineshape setting a single wavelet will suffice.

When the wavelet is a symmetric zero-mean function, such as the mexican hat (aka. Ricker or Marr) wavelet,

$$\psi_a(f) = \frac{2}{\sqrt{3a\pi^{1/4}}} \left(1 - \frac{f^2}{a^2}\right) \exp\left(-\frac{f^2}{2a^2}\right), \quad (2)$$

the convolution suppresses a locally smooth baseline. If the baseline can locally be modelled as a constant plus an odd signal, $b(f) = \delta + g(f)$ where $g(f) = g(-f)$, the convolution of the baseline with the wavelet is zero,

$$(b * \psi)(f) = \int_{-\infty}^{\infty} b(f') \psi_a(f' - f) df' = 0, \quad (3)$$

i.e. a locally smooth baseline does not influence the result. In practice, when the signal is measured in discretized frequency bins, the convolution is a discrete sum. Defining the discrete frequencies as $f = 1, 2, \dots, F$, and assuming the convolution kernel is defined on $f = 1, 2, \dots, W$ and centered at $(W + 1)/2$, we can write the convolution as

$$c[j] = \sum_{f=1}^W s[f + j] \psi_a[f], \quad (4)$$

where terms indexed by square brackets are discrete sequences. The convolution is devoid of issues with boundary conditions for $j = 0, 1, \dots, F - W$, and to simplify the discussion, we restrict the convolution to this range. Finally, the argument j^* that maximizes $c[j]$ is chosen, and the peak

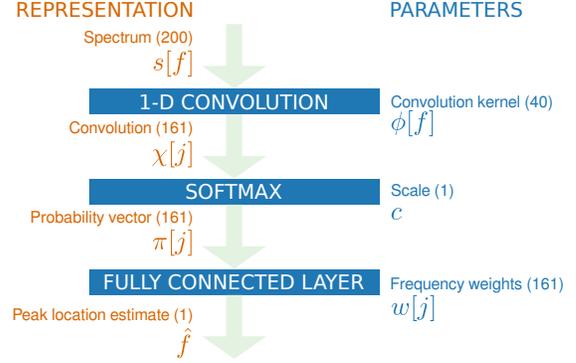


Fig. 1: Convolutional neural network architecture. Dimensions of input, output, intermediate data representations, and parameters used in experiments are indicated in parentheses.

location is estimated as $\hat{f} = j^* + (W + 1)/2$. To yield a sufficient frequency resolution beyond the given discrete resolution, the signal can be up-sampled before the convolution. The parameter(s) of the wavelet must be chosen in an appropriate manner. In our implementation we upsample by a factor of 100; use the mexican hat wavelet, and choose the wavelet width a selected by cross validation.

2.2. Formulation as a convolution neural network

We now move on to formulate the peak localization procedure as a trainable neural network. The first step is, as in the wavelet approach, a 1-dimensional convolution,

$$\chi[j] = \sum_{f=1}^W s[f + j] \phi[f], \quad j \in 1, 2, \dots, F - W. \quad (5)$$

Here, $s[f]$ is the measured spectrum and $\phi[f]$ are trainable kernel parameters. Rather than determine the location of the maximum of the convolution, we scale $\chi[j]$ by a trainable parameter c and pass it through a soft-max layer,

$$\pi[j] = \frac{\exp(c \cdot \chi[j])}{\sum_{k=0}^{F-W} \exp(c \cdot \chi[k])}. \quad (6)$$

Due to the soft-max operator, the sequence $\pi[j]$ is non-negative and sums to one, and can thus be thought of as a probability distribution over the discrete frequency bins that represents the possible locations of the peak. We finally have a linear readout layer,

$$\hat{f} = \sum_{j=0}^{F-W} \pi[j] w[j]. \quad (7)$$

If the weights in the readout layer are chosen as $w[j] = j + (W + 1)/2$, this layer will simply compute the expected value of the peak location according to the distribution $\pi[j]$. In practice, we let $w[j]$ be trainable parameters of the model.

The neural network architecture is illustrated in Fig. 1, which also shows the dimensionalities of all quantities used in the subsequent experiments.

2.3. Oracle benchmarks

We include for comparison two "oracle" benchmark peak localization procedures, in both of which the true baseline is removed from the signal.

Oracle peak picking: Here the peak localization is simply estimated by picking the frequency with the highest amplitude in the baseline corrected signal. This benchmark is included to demonstrate the performance of a procedure which does not model the peak lineshape.

Oracle convolution: In the case when there is no baseline and additive Gaussian noise, the least squares optimal procedure for estimating the peak location is to convolve the spectrum by the true peak lineshape and pick the maximum. In our implementation of this benchmark method, we up-sampled the spectrum by a factor of 100 to yield a frequency resolution beyond the discretization. This benchmark is included as an estimate of the best attainable performance.

3. EXPERIMENTAL EVALUATION

To demonstrate the performance of the proposed method, we carried out a cross validation study on a synthetic data set at different signal-to-noise ratios: We compared convolutional neural network (CNN) method with the continuous wavelet approach (using the mexican hat wavelet) as well as the two oracle baselines. In all experiments we used $F = 200$ frequency bins, and the width of the wavelets and convolution kernels were set to $W = 40$ to readily accommodate the width of the simulated peak lineshape.

3.1. Synthetic data generation

We generated a set of simulated spectra according to Eq. (1) with parameters chosen to mimic a spectroscopy scenario with a strong baseline and low signal-to-noise ratio.

Baseline: We generated a smoothly varying baseline by first generating a unit variance Gaussian random walk. The random walk was then filtered forward and backwards using a second order Butterworth lowpass filter with a normalized cutoff frequency of 0.01.

Peak: In each simulated spectrum we added a single Voigt-shaped peak,

$$v(f) = \frac{1}{\sigma\sqrt{2\pi}} \operatorname{Re} \left[w \left(\frac{f+i\gamma}{\sigma\sqrt{2}} \right) \right], \quad (8)$$

where w is the Faddeeva function and i is the imaginary unit. The peak shape parameters were fixed at $\sigma = \gamma = 1$ yielding a full width at half maximum of approximately 3.6. The peak was rescaled to have unit maximum magnitude. The peak locations were chosen uniformly at random in the interval $W \leq f \leq F - W$, so that no peaks occurred close to the boundaries of the spectra.

Noise: Independent Gaussian noise was added to each frequency bin. The noise was scaled by a factor s to achieve a

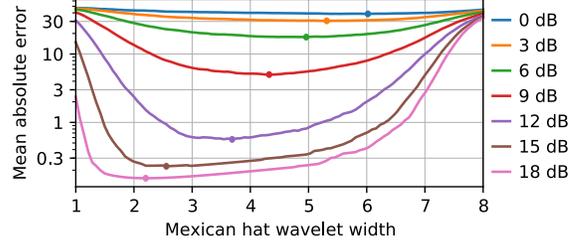


Fig. 2: Mean absolute error of peak localization for the continuous wavelet method for varying width of the mexican hat wavelet at different noise levels. The minima (indicated by dots) are used in the subsequent experiments.

desired peak signal-to-noise ratio (PSNR), computed as

$$\text{PSNR} = 10 \log_{10}(s^{-2}). \quad (9)$$

In our experiments we evaluated PSNR values from 0 dB to 18 dB in steps of 3 dB.

Crossvalidation: We generated a training, validation, and test set, each consisting of $N = 10\,000$ spectra. The training set was used to fit parameters in the neural network; the validation set was used for model selection for the wavelet methods; and the test set was used for final evaluation and comparison of the methods. Examples of a simulated spectrum at different PSNR are given in Fig. 4a.

3.2. Optimal wavelet width

In the continuous wavelet approach, to choose the optimal wavelet width, a , for the mexican hat wavelet, we computed the mean absolute error on the validation set for each PSNR for a range of wavelet widths ranging from 1 to 8.

Result: Fig. 2 shows the result of the analysis as well as the optimal wavelet width for each PSNR. For each given PSNR, the continuous wavelet method is not strongly sensitive to the choice of the width parameter, as none of the curves have a sharp minimum; however, the optimal wavelet width depends strongly on the PSNR, where low PSNR calls for a relatively wide wavelet, whereas high PSNR calls for a more narrow wavelet.

3.3. Performance evaluation

We fitted the parameters of the CNN on the training set using the L-BFGS optimization algorithm with the absolute error of the peak location as cost function,

$$\text{cost}(\phi, c, w) = \sum_{i=1}^N |f_i - \hat{f}_i|. \quad (10)$$

where i runs over training examples. The parameters were initialized as follows: The convolution kernel $\phi[f]$ was set to a mexican hat wavelet with optimal width chosen as for the

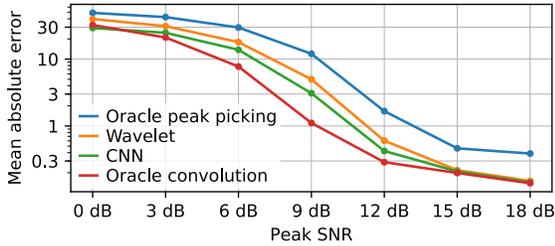


Fig. 3: Mean absolute error of peak localization estimators at different noise levels. The two benchmark “oracle” estimators were run with the ground truth baseline subtracted.

continuous wavelet method, the scale was set to $c = 10$, and the frequency weights were set to $w[j] = j + (W + 1)/2$.

Result: Fig. 3 shows the mean absolute error on the test set. The CNN consistently outperforms the wavelet approach except at the highest PSNR of 18 dB, where the performance of both methods coincide with the oracle convolution. At the lowest PSNR of 0 dB, the CNN performs better than the oracle convolution: We believe this happens because the oracle convolution is optimal in terms of fitting the signal in the least squares sense, but not optimal in estimating the peak location in terms of absolute error. Fig. 4a shows the estimated convolution kernel as well as the optimal width mexican hat wavelet used both in the continuous wavelet method and as initialization in the CNN. Similar to the optimal mexican hat wavelet, the learned kernel function appears slightly wider for low PSNR and narrower for high PSNR. For high PSNR values (15–18 dB) the estimated kernel is smooth and symmetric, whereas for lower PSNR the estimated kernel appears more “noisy”.

3.4. Learning curve

To assess how much data is needed to train the CNN method, we computed learning curves for the method by training on different number of spectra ranging from 100 to 10 000.

Result: Fig. 4b shows how the performance of the CNN method improves with increasing training set size in comparison with the wavelet approach. For low PSNR, 0–9 dB, the CNN outperforms the wavelet method even with only 100 training examples. At high PSNR, the CNN needs more training examples to outperform the wavelet method: We believe that a part of the explanation is that the CNN suffers somewhat from overfitting the training data at high PSNR, and this could possibly be remedied by suitable regularization.

4. CONCLUSION AND DISCUSSION

In this paper we have proposed a convolutive neural network approach to peak localization and baseline suppression. We have demonstrated that the proposed approach can leverage

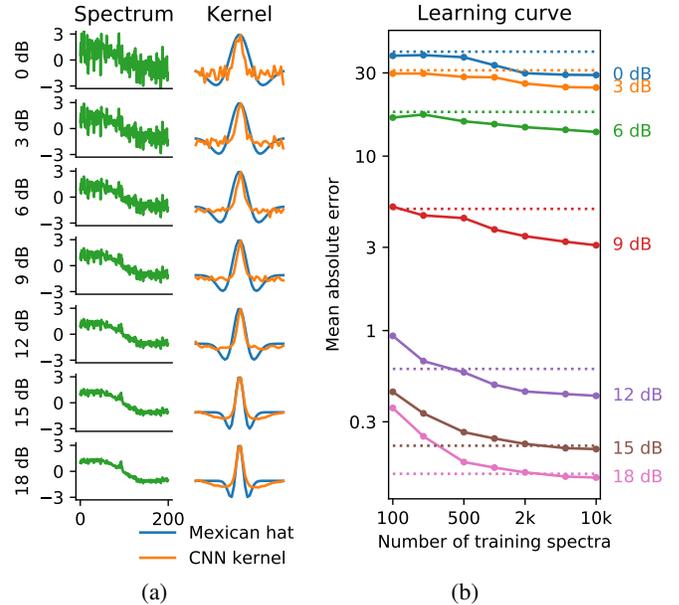


Fig. 4: (a) Examples of a spectrum at different PSNR levels, optimal mexican hat wavelet and the learned CNN kernel. (b) Learning curves for the convolutional neural network, showing the mean absolute error on the test set for different training set sizes, averaged over 5 re-runs. Dotted lines show results for the continuous wavelet method for comparison.

supervised training data to yield superior performance in comparison with the continuous wavelet approach, which we consider the state of the art unsupervised method.

To make the proposed method practically useful, several possible extensions should be considered. In the current formulation it was assumed that the spectral peak shape is constant. In practice, this might not be a valid assumption, and to handle this we envision extending the CNN by combining results from multiple learned kernel functions, which can model the span of possible lineshapes. Furthermore, it was assumed that the spectra always contained a single peak at fixed PSNR. In most practical applications, a more general assumption would be that there is an unknown number of peaks, and that the number of peaks should be inferred from the data. We envision that one way to handle this would be to include multiple peak location estimators and endow them with an attention mechanism so that each estimator will focus on a sub-range of frequencies.

With these future research directions in mind, we envision that the proposed CNN architecture can be of great value, incorporated in a larger spectral signal processing pipeline that is trained end-to-end.

5. REFERENCES

- [1] Derek J. Gardiner and Pierre R. Graves, Eds., *Practical Raman Spectroscopy*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1989.
- [2] Chao Yang, Zengyou He, and Weichuan Yu, “Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis,” *BMC Bioinformatics*, vol. 10, 2009.
- [3] Xiaoran Ning, Ivan W. Selesnick, and Laurent Duval, “Chromatogram baseline estimation and denoising using sparsity (BEADS),” *Chemometrics and Intelligent Laboratory Systems*, vol. 139, pp. 156–157, 2014.
- [4] J.P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M.B. Sandler, “A tutorial on onset detection in music signals,” *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1035–1047, sep 2005.
- [5] Philip Healy, Graham Hunt, Steven Kilroy, Theo Lynn, John P. Morrison, and Shankar Venkatagiri, “Evaluation of peak detection algorithms for social media event detection,” in *2015 10th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, nov 2015, pp. 1–9, IEEE.
- [6] Abraham Savitzky and Marcel E. J. Golay, “Smoothing and Differentiation of Data by Simplified Least Squares Procedures.,” *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, jul 1964.
- [7] Jiangtao Peng, Silong Peng, Qiong Xie, and Jiping Wei, “Baseline correction combined partial least squares algorithm and its application in on-line Fourier transform infrared quantitative analysis,” *Analytica Chimica Acta*, vol. 690, no. 2, pp. 162–168, apr 2011.
- [8] M.J. Shensa, “The discrete wavelet transform: wedding the a trous and Mallat algorithms,” *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464–2482, 1992.
- [9] S Sigurdsson, P.A. Philipsen, L.K. Hansen, J Larsen, M Gniadecka, and H.C. Wulf, “Detection of Skin Cancer by Classification of Raman Spectra,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 10, pp. 1784–1793, oct 2004.
- [10] Andreas F. Ruckstuhl, Mathew P. Jacobsen, Robert W. Field, and James A. Didd, “Baseline subtraction using robust local regression estimation,” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 68, no. 2, pp. 179–193, 2001.
- [11] Vincent Picaud, Jean-Francois Giovannelli, Caroline Truntzer, Jean-Philippe Charrier, Audrey Giremus, Pierre Grangeat, and Catherine Mercier, “Linear MALDI-ToF simultaneous spectrum deconvolution and baseline removal,” *BMC Bioinformatics*, vol. 19, no. 1, pp. 123, dec 2018.
- [12] M. Moores, K. Gracie, J. Carson, K. Faulds, D. Graham, and M. Girolami, “Bayesian modelling and quantification of Raman spectroscopy,” *ArXiv 1604.07299*, Apr. 2016.
- [13] P. Du, W. A. Kibbe, and S. M. Lin, “Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching,” *Bioinformatics*, vol. 22, no. 17, pp. 2059–2065, sep 2006.