

# Iterative peak-fitting of frequency-domain data via deep convolution neural networks

Seong-Heum Park<sup>a</sup>, Hyeong Seon Park<sup>a</sup>, Hyunbok Lee<sup>a,b,\*</sup>, Heung-Sik Kim<sup>a,b,\*</sup>

<sup>a</sup>*Department of Physics, Kangwon National University, Chuncheon 24341, Korea*

<sup>b</sup>*Institute for Accelerator Science, Kangwon National University, Chuncheon 24341, Korea*

## Abstract

High-throughput material screening for the discovery and design of novel functional materials requires automatized analyses of theoretical and experimental data. Here we study the subject of human-free analyses of one-dimensional spectroscopic data, *e.g.* in the frequency domain, via employing deep convolution neural network. Specifically, we trained various deep convolution neural network and benchmarked their performance in decomposing one-dimensional noisy data into multiple nonorthogonal peaks in an iterative manner, after which a conventional basin-hopping algorithm was applied to further reduce residual fitting error. Among six different network architectures, a variant of “Squeeze-and-excitation” network (SENet) structure that we first propose in this study shows the best performance. Dependency of training performance with respect to the choice of the loss function is also discussed. We conclude by applying our modified SENet model to experimental photoemission spectra of graphene, MoS<sub>2</sub>, and WS<sub>2</sub> and address its potential applications and limitations.

**Keywords:** Deep learning, convolution neural network, peak fitting

**2010 MSC:** 00-01, 99-00

## 1. Introduction

Thanks to recent advancements in the capabilities of synthesizing, characterizing, and analyzing new materials, there has been a surge of theoretical and experimental data on potentially interesting systems that has accumulated over the last decade. Specifically, developments in *ab initio* electronic structure calculation methods such as density functional theory[1, 2] have enabled the compilation of extensive computational materials databases such as MATERIALS PROJECT<sup>1</sup>, OPEN QUANTUM MATERIALS DATABASE (OQMD)<sup>2</sup>, or NOVEL MATERIALS DISCOVERY (NOMAD)<sup>3</sup>, based on which data mining and machine-learning approaches can be applied to establish statistical models with enough predicting power for novel materials discovery and inverse materials design[3, 4, 5, 6, 7, 8, 9].

\*Corresponding authors

Email addresses: hyunbok@kangwon.ac.kr (Hyunbok Lee), heungsikim@kangwon.ac.kr (Heung-Sik Kim)

<sup>1</sup><https://materialsproject.org>

<sup>2</sup><http://oqmd.org>

<sup>3</sup><https://nomad-coe.eu>

A relatively under-explored area, compared to analyzing and learning from cumulated computational data for training predictive models, is applying machine learning for analyses of experimental spectroscopic data. There have been a few earlier works on the application of machine learning in distilling useful physical information from experimental photoemission and scanning tunneling microscopy spectra[10, 11, 12], and specifically a recent study on application of deep neural network to core-level x-ray photoemission spectra[11] showed a qualitative agreement with experimental results in quantifying stoichiometry of chemical elements of given test compounds, but with about 10% error in actual composition ratio. Compared to its importance in condensed matter and materials sciences studies, however, the number of subsequent studies has been rather limited. Several difficulties in analyzing experimental spectroscopic data are *i*) inherent non-uniqueness in decomposing spectra into non-orthogonal elements (such as peaks), *ii*) limited amount of experimental spectra which are often insufficient to train deep neural networks, and *iii*) superiority of human recognition based on physical intuitions over machine-trained models in traditional spectroscopy studies.

Since there are a number of spectroscopic tools where the data is measured in energy or frequency domain, such as, not limited to, Raman, photoemission, light absorption, photoluminescence, nuclear magnetic resonance, and inelastic x-ray or neutron scatterings, computational tools to automatize analyses of experimental spectra are essential in order to enable data-driven materials scientific studies. For this purpose, here we combined deep convolution neural networks (CNN) and conventional basin-hopping optimization techniques to decompose a given spectrum  $F(\omega)$  in one-dimensional parameter ( $\omega$ ) domain into multiple peak functions  $f_i(\omega)$  ( $i$  being a peak label), so that the error between original  $F(\omega)$  and the sum of  $f_i(\omega)$  becomes minimized with minimal human intervention. Specifically, in search of the best CNN model to tackle the non-unique fitting problem, we employed six deep CNN architectures with varying network topology and complexity; 5 of them have been recently suggested, while the last one (modified SENet, denoted as m-SENet hereafter) is newly proposed in this work. We found that, among the six proposed CNN architectures employed, our m-SENet shows the best fitting performance.

Our trained CNN models were then used to find a peak with the largest underlying area so that the fitted peak can be subtracted repeatedly until no peak feature is left in the spectrum; see an illustration of how this works in Fig. 1, where the trained CNN almost exactly captures peak features. Limitation in the size of accessible experimental spectroscopy data was circumvented by using synthetic spectra generated via the pseudo-Voigt function[13]. Finally, errors that occur in the regressions process are corrected after all peaks are extracted by applying conventional global optimizations such as basin-hopping algorithm, which fine-tunes CNN-extracted peak parameters and minimizes the loss function. After checking the fitting performance of our model, we applied our best trained model to actual experimental photoemission data from graphene, MoS<sub>2</sub>, and WS<sub>2</sub> and confirmed the predicting power of our approach in fitting the actual experimental spectroscopy data. We conclude by discussing potential loopholes of our approach and necessary improvements.

## 2. Methods

### 2.1. Description of neural network architectures

As mentioned in the Introduction section, we need a well-trained CNN architecture to accurately subtract a peak with the largest underlying area from a spectrum. For this purpose, we benchmarked fitting performances of six convolution neural network (CNN) architectures; i) LeNet

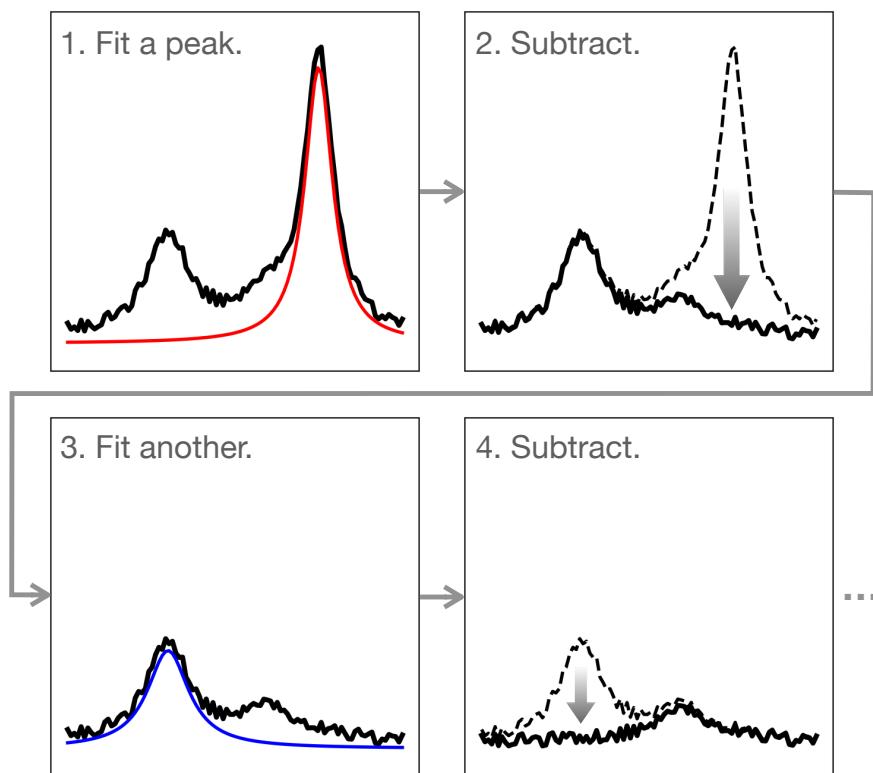


Figure 1: A cartoon illustration of decomposing a noisy spectrum (black curve) with multiple peak features by sequentially applying regression and subtraction of the largest-area peak (red and blue curves) until no peak feature is left.

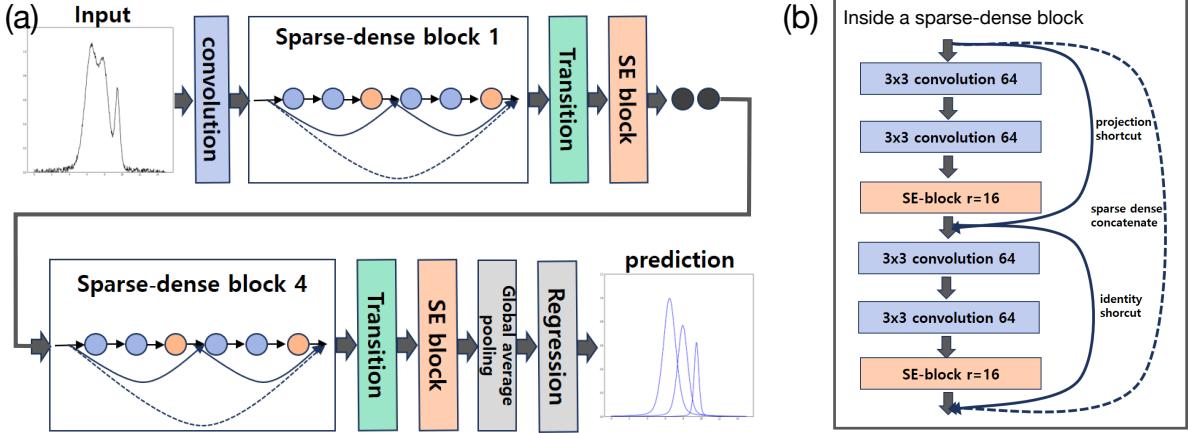


Figure 2: (a) A schematic representation of m-SENet architecture, where detailed layering structure inside a sparse-dense block is depicted in (b).

[14, 15], Alex-ZFNet [16, 17], VGGNet [18], ResNet [19], SENet [20], and a modified version of SENet (hereafter denoted as m-SENet), listed in order of increasing complexity in network structure. LeNet is the first proposed CNN architecture, which consists of alternating convolution and subsampling (pooling) layers. Alex-ZFNet[16, 17] can be considered an improvement of LeNet by introducing maxpooling, dropout, and rectified linear unit (ReLU) activation function. VGGNet[18] takes a different path by adopting a uniform network structure but with increased number of convolution filters. ResNet[19] employs the so-called ‘skip connections’ to enhance back-propagation of the loss function gradient deep into the network layers. In SENet[20] a so-called ‘squeeze-and-excitation (SE) block’ is introduced to improve performance with a small amount of computational cost increase.

For this work we employed the m-SENet (see Fig. 2), where the idea of SE block from SENet was implemented into the so-called ‘sparse-dense block’ as shown in Fig. 2(a) and (b). Each sparse-dense block consists of serially connected pairs of convolution layers (CONV) and a SE block, where identity mapping connects input and output layers of two CONV-CONV-SE triplets in a sparse-dense block. Batch normalization and SE layers were placed between adjacent sparse-dense blocks to rebalance weights of convolution feature maps. This structure allows efficient learning even in extremely deep models. Finally, the regression layer was attached through global average pooling. The number of total trainable parameters was 4,183,830 and ‘overlapped average pooling’ was used in all subsampling processes.

## 2.2. Generation of training, validation, and test datasets

Compared to the large amount of dataset required to train deep neural networks, the amount of accumulated spectroscopic data from experimental measurements is often insufficient. Hence we generated a synthetic dataset for training networks employing pseudo-Voigt profile[13], which is often adopted for fitting peaks from experimentally measured x-ray diffraction or photoemission data. The following form was chosen to generate each peak in the artificial dataset,

$$f(\omega; \omega_0, I_0, \delta) = I_0 \left( I_G e^{-\frac{\log(2)(\omega-\omega^2)}{\delta^2 \omega_G^2}} + I_L \frac{1}{1 + \frac{(\omega-\omega_0)^2}{\delta^2 \omega_L^2}} \right), \quad (1)$$

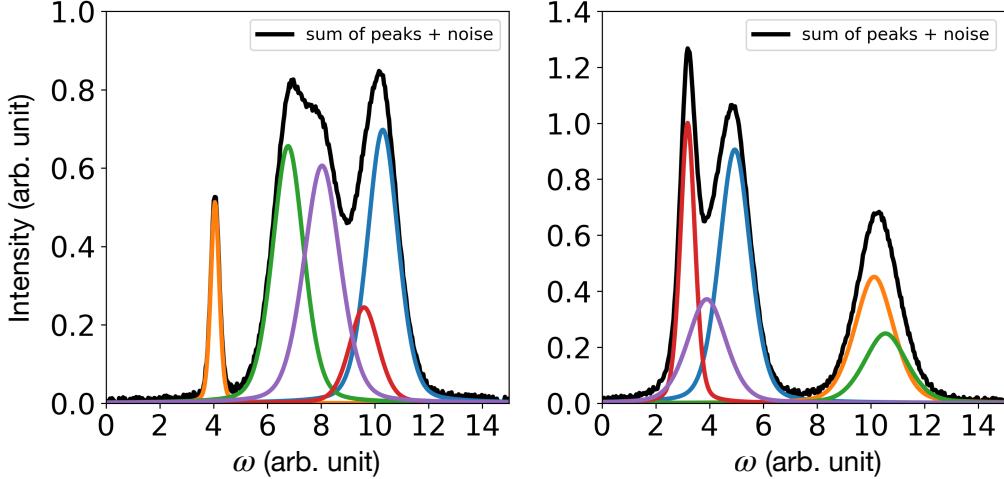


Figure 3: Two examples of synthetically generated data with multiple (up to 5) pseudo-Voigt profiles. Colored and black lines depict each pseudo-Voigt peaks and addition of all peaks with gaussian noise, respectively.

where  $\omega$  can be considered frequency (energy) in spectroscopic data, with  $\omega_0$ ,  $I_0$ , and  $\delta$  being randomly generated position, maximum height, and (dimensionless) width of the peak, respectively.<sup>4</sup>  $\omega_G$ ,  $\omega_L$ ,  $I_G$ , and  $I_L$  are set to be 0.510, 0.441, 0.7, and 0.3, respectively; this choice of parameters in the above pseudo-Voigt profile was adopted to fit experimental spectra in a previous photoemission study[21].

In each one-dimensional synthetic data up to 5 random pseudo-Voigt functions were added, with the addition of gaussian noise of which standard deviation being 2% of the maximum allowed  $I_0$  (see Fig. 3). We also enforced centers of neighboring peaks to be no closer than 20% of the maximum allowed  $\delta$ . Background signals other than gaussian noises are not considered in this work, however, it was previously reported that filtering monotonically increasing or decreasing background signals can be easily done employing a simple CNN network.[22]

### 2.3. Implementation, train, and validation of neural networks

To implement various CNN structures and train them we employed KERAS[23] framework and TENSORFLOW[24] backend. Training of CNNs was done using a Nvidia RTX 2080Ti graphic processing unit. Four different target variables — center  $\omega_0$ , width  $\delta$ , and amplitude  $I_0$  of the peak with the largest underlying area, and the number of peaks in the spectrum — contribute to the total loss function, which is the sum of loss functions for center, width, amplitude, and peak numbers. To regularize contributions of different loss components in the total loss, weights of 1, 10, 20, and 2 were assigned to each part of the loss function optimizing  $\omega_0$ ,  $\delta$ ,  $I_0$ , and peak number, respectively. ADAM[25] optimizer was employed throughout the whole training and test jobs for this study. The stochastic gradient descent (SGD) method[26] was tested in comparison to ADAM results and found to show almost the same optimization performance compared to ADAM in this study. In all cases, LeakyReLU activation function was used[27], and trainings were done during 50 epochs with 512 batch-size. The learning rate was reduced to 10% after 20th and 40th epochs to avoid overfitting.

---

<sup>4</sup> $\omega_0$ ,  $I_0$ , and  $\delta$  are randomly chosen in the range of  $0.13\omega_{\max} < \omega_0 < 0.87\omega_{\max}$ ,  $0.02\omega_{\max} < \delta < 0.13\omega_{\max}$ , and  $0.05 < I_0 < 1.05$ , respectively, where  $0 \leq \omega \leq \omega_{\max}$  is our frequency domain.

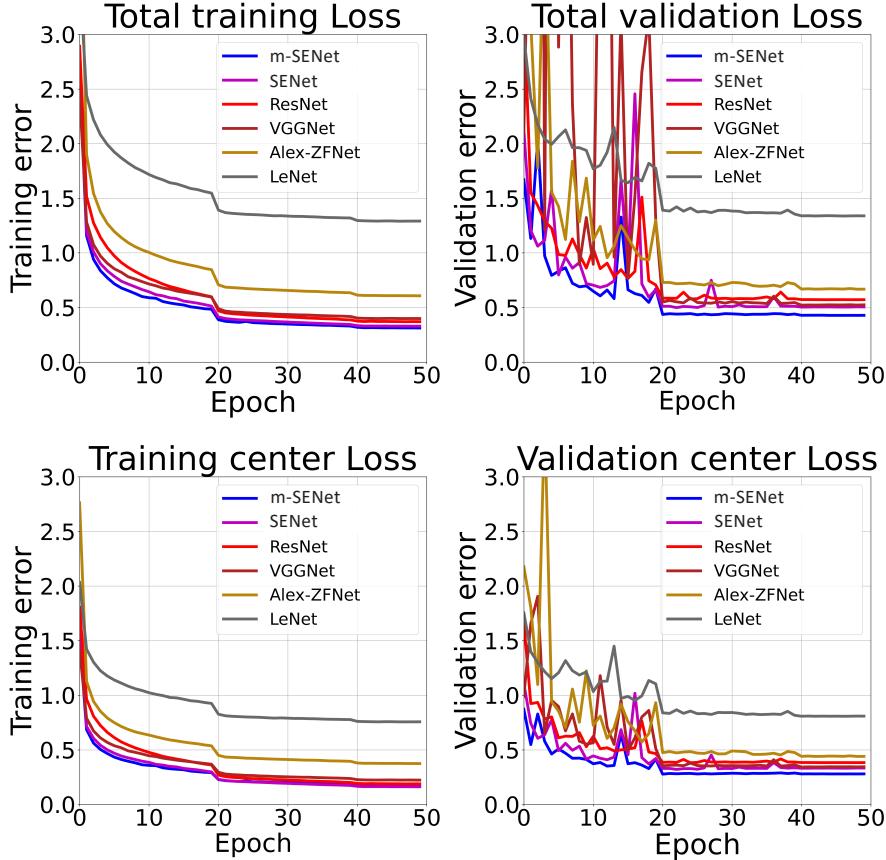


Figure 4: Total (upper row) and peak center loss functions (lower row) of six different CNN architectures, with respect to training epochs. Note that learning rate was reduced to 10% after 20th and 40th epochs. Mean-square error loss function was employed. Left and right columns show training and validation losses, respectively.

No dropout layer was used in this study.

For the CNN benchmark results presented in this work, we generated  $1.5 \times 10^6$  synthetic spectra, among which  $1.2 \times 10^6$ ,  $1.5 \times 10^5$ ,  $1.5 \times 10^5$  data were used for training model, validating and testing model performance, respectively. Later we increased the size of the dataset to  $10^7$  to check performance improvements.

### 3. Results

#### 3.1. CNN training and validation

Fig. 4 shows a summary of training and validation mean-square losses from different CNN architectures (SGD as optimizer). Among the four target properties (number of peaks in each spectrum and center, width, and amplitude of maximum area peak) peak center loss, which tends to yield the largest error, were plotted as well. A gradual reduction of both the training and validation loss as we go from LeNet to m-SENet is evident, showing better representational power of our m-SENet model in peak detection and fitting problem compared to other CNN architectures.

Note that training and validation results shown in Fig. 4 were obtained using mean-square-error (MSE) as the loss function, which is a widely adopted choice. In several cases, however, MSE tends to overestimate the effects of very few outliers in our randomly generated dataset and may

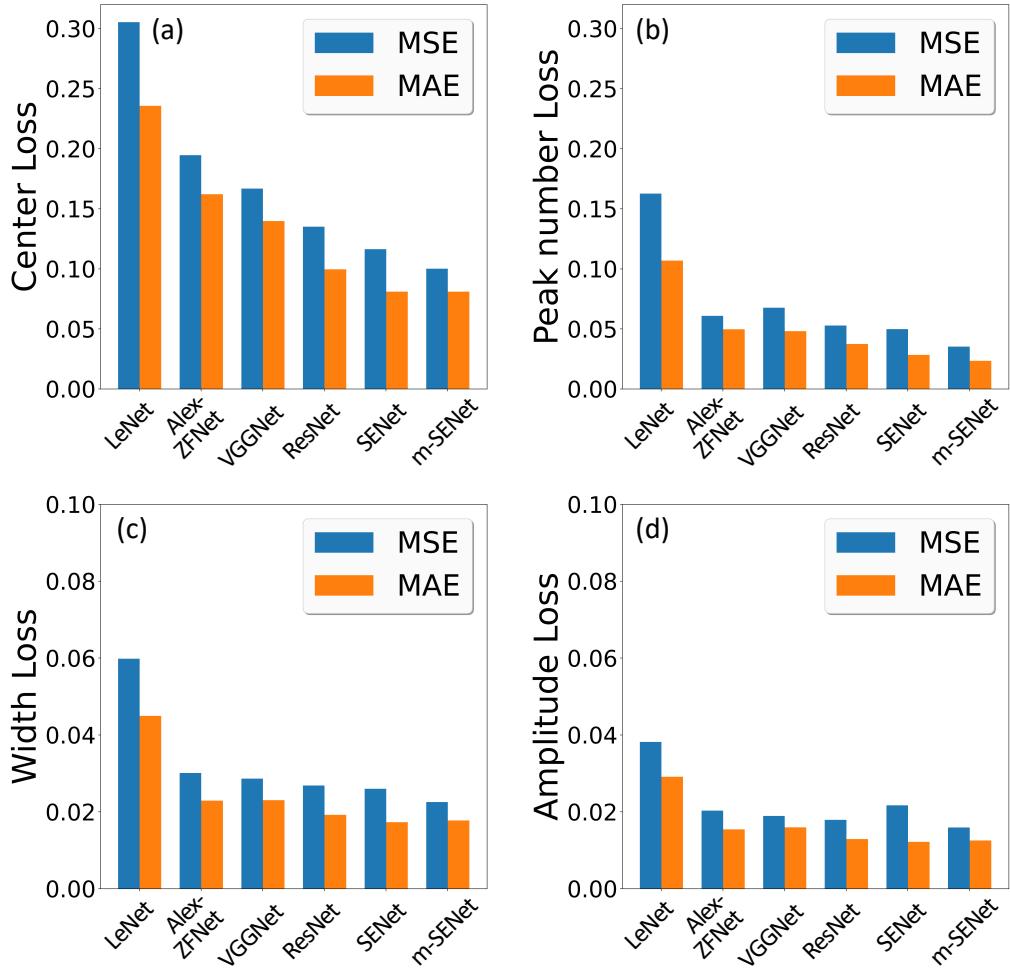


Figure 5: MAE loss functions — (a) center, (b) number of peaks, (c) width, and (d) amplitude of the largest-area peak — of six different CNN architectures on the same test dataset. Blue and orange columns show results employing MSE and MAE losses during the network training and validation, respectively.

Table 1: Comparison between MAE loss functions from m-SENet trained with 1.5 and 10 millions datasets, where the values were obtained by applying same test dataset to both trained models.

	center	width	amplitude	number of peaks
1.5M set	0.083	0.018	0.013	0.023
10M set	0.056	0.012	0.009	0.013

lead to poor predicting performance of the model. In such cases, mean-absolute-error (MAE) can be a better option. Interestingly, our multiple peak fitting problem was found to be such one. To compare effects of MSE and MAE loss functions in model training and validation, we separately trained each CNN model using MSE and MAE, then applied each trained model in fitting the test dataset and computed MAE losses for comparison in equal footing. Fig. 5 summarizes the results, where test losses of all four target variables (number of peaks, and center, width, and amplitude of the largest-area peak) in all CNN architectures show lower value when MAE was used in model training. From the result, it can be seen that SENet and m-SENet with MAE training loss function show almost the same fitting performance in this study.

Once we confirmed that m-SENet shows the best fitting performance, we increased the size of the total dataset from 1.5 million to 10 million to enhance the performance of the model, because accurate peak fitting is crucial in the sequential subtraction of peaks from the spectrum. Table 1 compares loss function results from m-SENet trained with 1.5 and 10 million datasets, which shows clear improvements employing a large training dataset. Note that we haven't tested saturation of model performance as a function of the size of the training dataset due to the limitation of our computational resources.

### 3.2. Iterative peak subtraction via trained CNN and basin-hopping optimization

After an accurate CNN model is prepared, it can be applied to sequentially subtract peak features from any given spectrum. Fig. 6 illustrates our iterative peak fitting and subtraction scheme. A potential issue in this type of feature subtraction is, even very small errors in peak fitting may cause wiggly and negative features in the spectrum after subtraction which may be mistaken as some meaningful peaks by the CNN. This issue can be circumvented by i) accurately determining the number of total peaks at the initial fitting and performing subtractions exactly the number of times and ii) truncating negative intensity after subtraction (see lower left panel of Fig. 6). Because the accuracy of peak number prediction is quite high (check Fig. 5 and Table 1), this should in most cases prevent the overall workflow from obtaining additional spurious peaks.

Lastly, after extracting all peak features in a given spectrum, conventional global optimization methods such as basin-hopping algorithms[28] can be applied to further optimize fitting results and reduce any residual error to minimum. Basin-hopping can be considered an improved simulated annealing method, which involves in an additional local minimization stage before the accept/reject decision at each iteration[28]. We used `scipy.optimize.basinhopping` function implemented in the SCIPY package<sup>5</sup>. Fig. 7 show an example optimization result on a couple of test data, where in Fig. 7(a) the iterative CNN application already resulted in fairly good agreement with the target spectrum. In another case (Fig. 7(b)), where CNN yields relatively poor agreement with

---

<sup>5</sup><https://www.scipy.org>

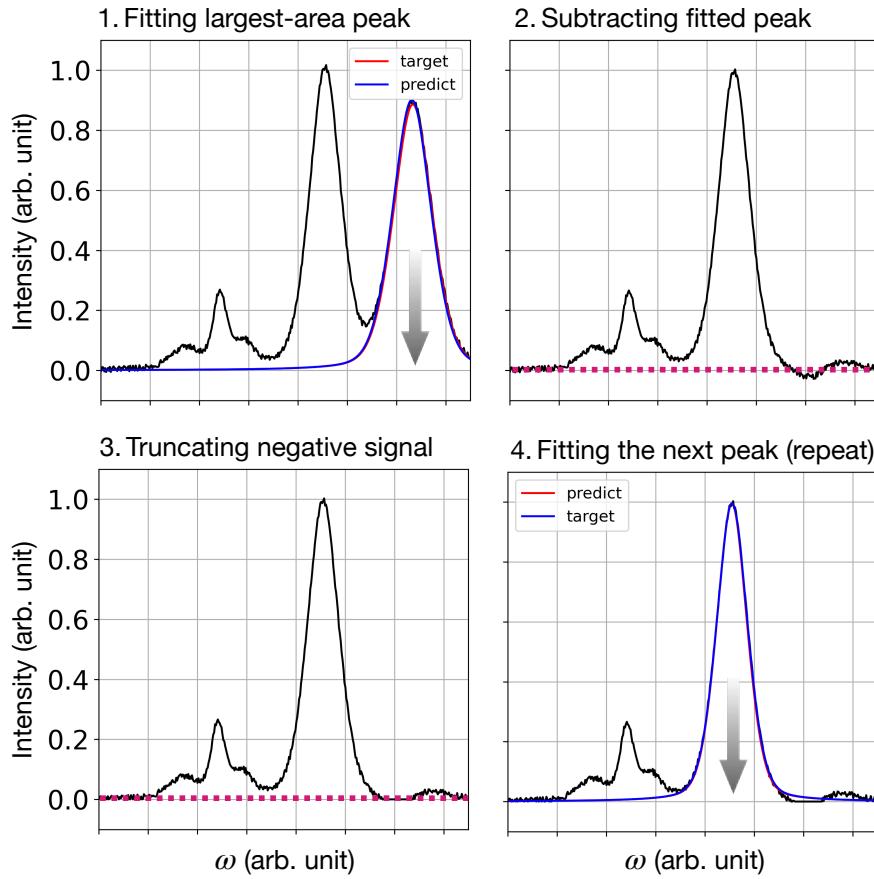


Figure 6: An illustration of iterative peak fitting and subtraction scheme we employed in this study. In the upper left and lower right panels, red and blue curves depict CNN-fitted and exact peak features respectively. Negative intensity caused by subtraction of inaccurate peaks (upper right panel, below the red dashed line) is truncated before applying subsequent peak fitting and subtraction.

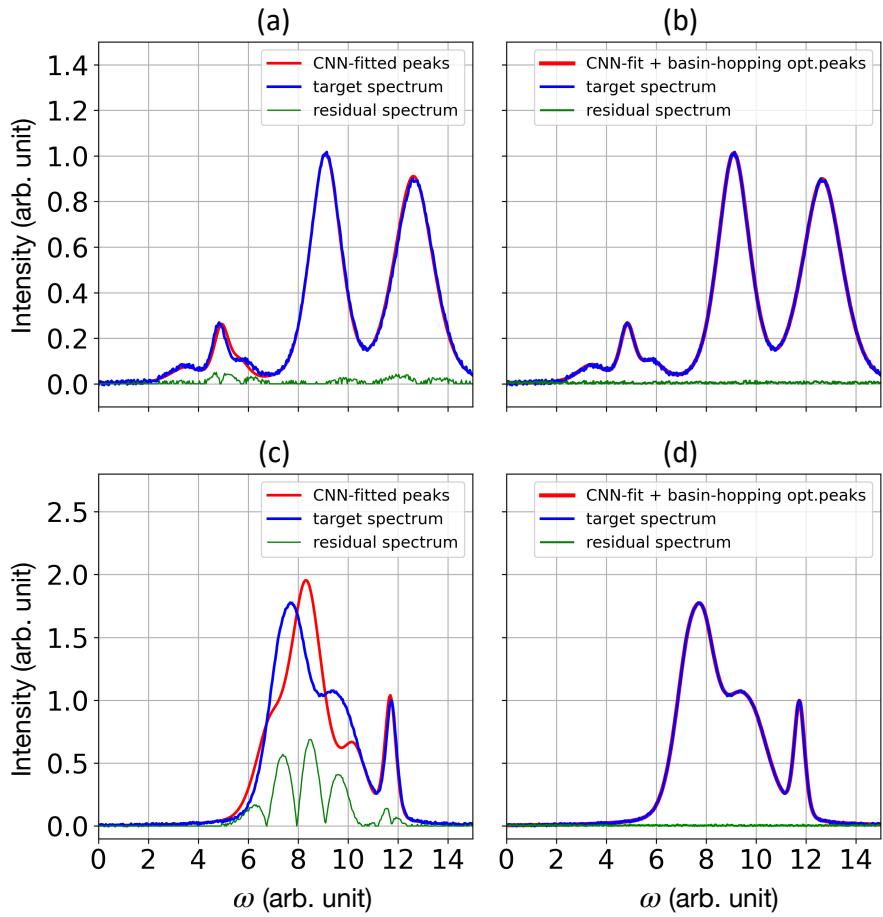


Figure 7: Comparison between peak fitting results before and after applying basin-hopping algorithm. Top (a,b) and bottom (c,d) rows show two different cases where the initial iterative CNN-fitting yield good and relative poor agreements with the target spectra, respectively, after which application of basin-hopping algorithm reduces any residual errors to almost zero.

Table 2: Comparison of human-fitted and machine-fitted peak parameters from experimental photoemission spectra as shown in Fig. 8.

	Human-fitted			Machine-fitted		
	center (eV)	width (eV)	Norm. amp.	center (eV)	width (eV)	Norm. amp.
Graphene, C 1s						
Peak 1	284.60	1.00	0.96	284.50	1.08	0.81
Peak 2	285.50	1.00	0.21	285.15	1.72	0.33
Peak 3	286.60	1.00	0.06	286.71	0.97	0.06
Peak 4	-	-	-	284.16	0.08	0.06
MoS <sub>2</sub> on Si, Mo 3d						
Peak 1	226.98	2.06	0.15	226.94	2.08	0.16
Peak 2	229.78	0.80	0.92	229.86	0.68	0.88
Peak 3	232.93	0.80	0.57	232.97	0.92	0.55
Peak 4	-	-	-	229.46	0.99	0.19
Peak 5	-	-	-	232.20	0.68	0.06
MoS <sub>2</sub> on Si, S 2p						
Peak 1	162.68	0.70	0.97	162.66	0.69	0.74
Peak 2	163.83	0.70	0.47	163.83	0.72	0.49
Peak 3	168.13	4.56	0.23	168.01	4.56	0.24
Peak 4	-	-	-	162.61	1.08	0.24
Peak 5	-	-	-	171.52	2.34	0.03
WS <sub>2</sub> on Si, S 2p						
Peak 1	163.13	0.70	0.91	163.15	0.74	0.95
Peak 2	164.33	0.70	0.44	164.32	0.68	0.48
Peak 3	168.18	4.46	0.24	168.19	4.35	0.24
Peak 4	-	-	-	166.21	2.21	0.03
Peak 5	-	-	-	171.48	2.78	0.03
WS <sub>2</sub> on Si, W 4f						
Peak 1	33.53	0.60	1.00	33.58	0.53	0.93
Peak 2	35.68	0.60	0.73	35.68	0.60	0.71
Peak 3	39.03	1.52	0.05	39.03	1.57	0.05
Peak 4	-	-	-	33.13	0.82	0.15
Peak 5	-	-	-	35.18	0.52	0.70

the target spectrum, it can be checked that basin-hopping algorithm remedies the situation starting from the initial guess of peak parameters provided by the CNN model.

#### 4. Application to experimental photoemission spectra

Finally, we apply our fitting method to experimental x-ray photoemission spectra of graphene, MoS<sub>2</sub>, and WS<sub>2</sub>, and the results are summarized in Fig. 8, where the top (Fig. 8(a-e)) and bottom (Fig. 8(g-k)) rows show human- and machine-fitted results, respectively. Note that background signals were subtracted from the raw data via IGOR PRO package, and that defect-induced shoulder-like features in Mo 3d (orange- and green-shaded area in Fig. 8(b)), Si 2s (green-shaded area in Fig. 8(c,d)), and W 4f (blue- and orange-shades in Fig. 8(e)) were intentionally ignored for the comparison with our machine-fitted results (see the residual signals therein, depicted as green curves).

It can be seen that, in all cases, our machine-fitted results yield better agreement with the raw data compared to human-fitted results (*i.e.* smaller residual signal) and our previous result[29].

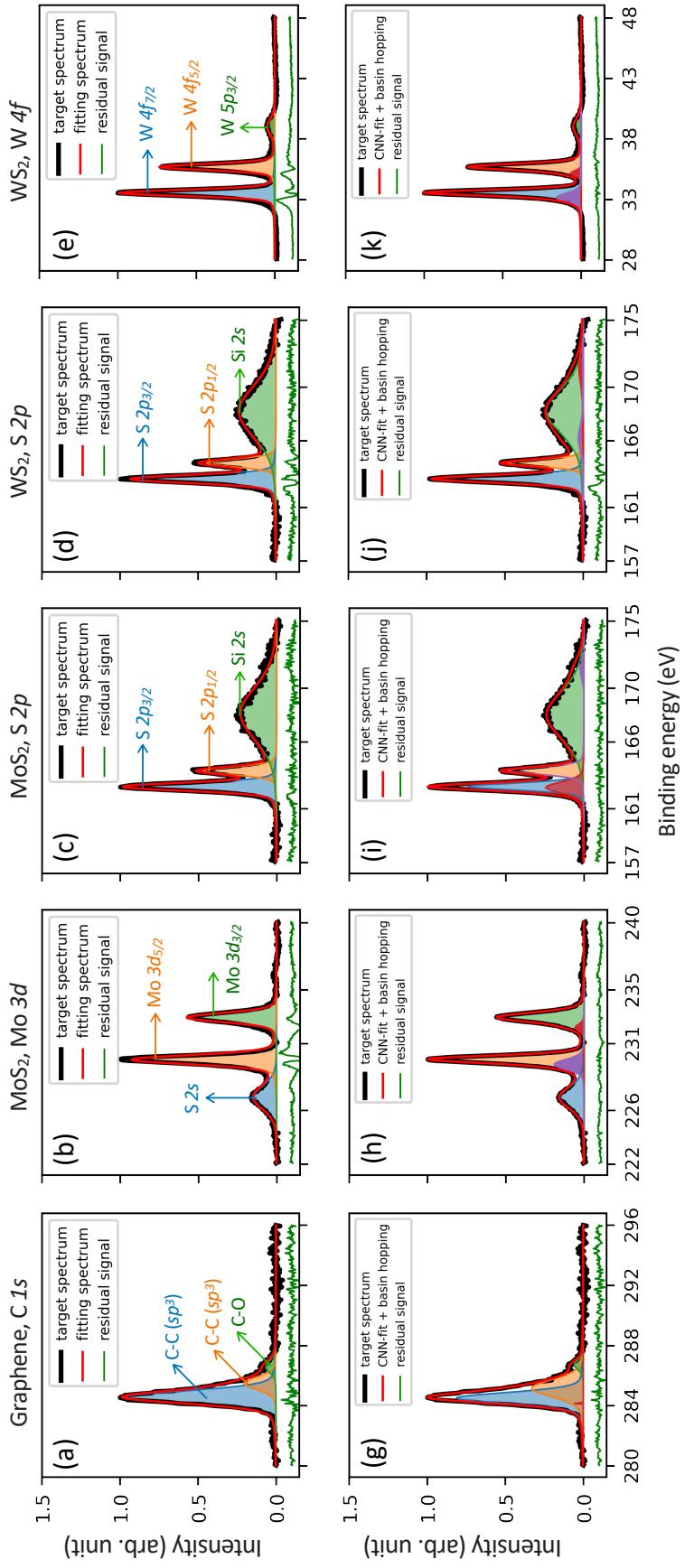


Figure 8: Human- (upper row) and machine-fitted (lower row) photoemission spectra of (a) graphene C 1s, (b) MoS<sub>2</sub> Mo 3d, (c) MoS<sub>2</sub> S 2p, (d) WS<sub>2</sub> S 2p, and (e) WS<sub>2</sub> W 4f. In each panel black, red, and green curves represent raw data with background signal subtracted, fitted spectrum, and residual signal, respectively, where fitted peaks are depicted as colored filled curves.

Note that in the human-fitted results, fitting for Mo 3d, Si 2s, and W 4f peaks result in noticeable residual signals, signifying the defected-induced shoulder-like features. Indeed our iterative fitting model predicts the presence of additional peaks in those cases (darker red and violet shades in Fig. 8(h,j,k), probably capturing overlooked signatures in human-fitting.

On the other hand, note that our fitting model generates spurious additional peaks in C-C  $sp^3$  in graphene and S 2p in MoS<sub>2</sub> (red shades in Fig. 8(g) and (i), respectively), which may be attributed to the choice of pseudo-Voigt function with fixed internal parameters in Eq. 1. The use of true Voigt function with variable Lorentzian-Gaussian ratio to generate training dataset may resolve this issue.

## 5. Discussion and conclusion

We comment that, in the generation of the training dataset, we did not incorporate knowledge on physical processes that generate broadening of peaks, and considered peak width as a uniform random parameter in a range of our choice. Since excitations located close in energy should undergo similar scattering processes, most of the experimental photoemission peaks in a given energy range often show similar peak widths. Consideration of such observation in data generation may enhance the fitting performance of our model. Also, the employment of the true Voigt function instead of the pseudo-Voigt function that we used may improve the performance even further, which may reduce the occurrence of spurious peaks in the fitting.

Overall, in this study, we've explored the potential power of CNN in analyzing spectroscopy data in the frequency domain and achieved significant improvement in fitting performance compared to our previous work[29] via the employment of advanced deep CNN architectures and the application of iterative peak subtraction. We also benchmarked the fitting performance of various deep CNN networks and confirmed that the m-SENet that we propose in this study yields the best performance. With the help of conventional global optimization techniques, our method produces comparable or sometimes even better fitting performance compared to human fitting, and may potentially be applicable to high-throughput analyses of spectroscopic data with minimal human intervention.

## 6. Acknowledgements

HSK acknowledges the support of the National Research Foundation of Korea (Basic Science Research Program, Grant No. 2020R1C1C1005900), and also thanks Hunpyo Lee, Juyong Lee, Hongkee Yoon, and Sooran Kim for fruitful discussions and insightful comments on the manuscript.

## References

- [1] R. G. Parr, [Density functional theory](#), Ann. Rev. Phys. Chem. 34 (1) (1983) 631–656.  
[arXiv:<https://doi.org/10.1146/annurev.pc.34.100183.003215>](https://doi.org/10.1146/annurev.pc.34.100183.003215), doi:[10.1146/annurev.pc.34.100183.003215](https://doi.org/10.1146/annurev.pc.34.100183.003215).  
URL <https://doi.org/10.1146/annurev.pc.34.100183.003215>
- [2] K. Burke, [Perspective on density functional theory](#), J. Chem. Phys. 136 (15) (2012) 150901.  
[arXiv:<https://doi.org/10.1063/1.4704546>](https://doi.org/10.1063/1.4704546), doi:[10.1063/1.4704546](https://doi.org/10.1063/1.4704546).  
URL <https://doi.org/10.1063/1.4704546>

- [3] A. Karpatne, G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, V. Kumar, Theory-guided data science: A new paradigm for scientific discovery from data, *IEEE Trans. Knowl. Data. Eng.* 29 (10) (2017) 2318–2331.
- [4] T. Xie, J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, *Phys. Rev. Lett.* 120 (2018) 145301.
- [5] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, *Nature* 559 (7715) (2018) 547–555. [doi:10.1038/s41586-018-0337-2](https://doi.org/10.1038/s41586-018-0337-2).
- [6] A. Zunger, Inverse design in search of materials with target functionalities, *Nat. Rev. Chem.* 2 (4) (2018) 0121. [doi:10.1038/s41570-018-0121](https://doi.org/10.1038/s41570-018-0121).
- [7] B. Sanchez-Lengeling, A. Aspuru-Guzik, Inverse molecular design using machine learning: Generative models for matter engineering, *Science* 361 (6400) (2018) 360–365. [doi:10.1126/science.aat2663](https://doi.org/10.1126/science.aat2663).
- [8] G. R. Schleider, A. C. M. Padilha, C. M. Acosta, M. Costa, A. Fazzio, [From DFT to machine learning: recent approaches to materials science—a review](#), *J. Phys.: Mater.* 2 (3) (2019) 032001. [doi:10.1088/2515-7639/ab084b](https://doi.org/10.1088/2515-7639/ab084b).  
URL <https://doi.org/10.1088/2515-7639/ab084b>
- [9] L. Himanen, A. Geurts, A. S. Foster, P. Rinke, [Data-driven materials science: Status, challenges, and perspectives](#), *Adv. Sci.* 6 (21) (2019) 1900808. [arXiv:https://onlinelibrary.wiley.com/doi/10.1002/advs.201900808](https://onlinelibrary.wiley.com/doi/10.1002/advs.201900808), [doi:https://doi.org/10.1002/advs.201900808](https://doi.org/10.1002/advs.201900808).  
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/advs.201900808>
- [10] Y. Zhang, A. Mesaros, K. Fujita, S. D. Edkins, M. H. Hamidian, K. Ch'ng, H. Eisaki, S. Uchida, J. C. S. Davis, E. Khatami, E.-A. Kim, [Machine learning in electronic-quantum-matter imaging experiments](#), *Nature* 570 (7762) (2019) 484–490. [doi:10.1038/s41586-019-1319-8](https://doi.org/10.1038/s41586-019-1319-8).  
URL <https://doi.org/10.1038/s41586-019-1319-8>
- [11] G. Drera, C. M. Kropf, L. Sangaletti, [Deep neural network for x-ray photoelectron spectroscopy data analysis](#), *Mach. Learn.: Sci. Technol.* 1 (1) (2020) 015008. [doi:10.1088/2632-2153/ab5da6](https://doi.org/10.1088/2632-2153/ab5da6).  
URL <https://doi.org/10.1088%2F2632-2153%2Fab5da6>
- [12] Y. Yamaji, T. Yoshida, A. Fujimori, M. Imada, [Hidden self-energies as origin of cuprate superconductivity revealed by machine learning](#) (2019).  
URL <http://arxiv.org/abs/1903.08060>
- [13] T. Ida, M. Ando, H. Toraya, Extended pseudo-Voigt function for approximating the Voigt profile, *J. Appl. Crystallogr.* 33 (6) (2000) 1311–1316. [doi:10.1107/S0021889800010219](https://doi.org/10.1107/S0021889800010219).

- [14] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
- [15] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural Comput. 1 (4) (1989) 541–551.
- [16] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012) 1097–1105.
- [17] M. D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: European conference on computer vision, Springer, 2014, pp. 818–833.
- [18] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).
- [19] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- [20] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.
- [21] H. Kim, H. Lee, H. Lee, Fabrication of poly(3-hexylthiophene-2,5-diyl) films with electro-spray deposition method, Jpn. J. Appl. Phys. 57 (7) (2018) 071601. [doi:10.7567/jjap.57.071601](https://doi.org/10.7567/jjap.57.071601).
- [22] M. N. Schmidt, T. S. Alstrøm, M. Svendstorp, J. Larsen, Peak detection and baseline correction using a convolutional neural network, in: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 2757–2761.
- [23] F. Chollet, et al., Keras, <https://keras.io> (2015).
- [24] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, [TensorFlow: Large-scale machine learning on heterogeneous systems](#), software available from tensorflow.org (2015).  
URL <https://www.tensorflow.org/>
- [25] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
- [26] L. Bottou, Large-scale machine learning with stochastic gradient descent, in: Y. Lechevallier, G. Saporta (Eds.), Proceedings of COMPSTAT'2010, Physica-Verlag HD, Heidelberg, 2010, pp. 177–186.
- [27] B. Xu, N. Wang, T. Chen, M. Li, Empirical evaluation of rectified activations in convolutional network, arXiv preprint arXiv:1505.00853 (2015).

- [28] D. J. Wales, J. P. K. Doye, [Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms](#), *J. Phys. Chem. A* 101 (28) (1997) 5111–5116. doi:[10.1021/jp970984n](https://doi.org/10.1021/jp970984n).  
URL <https://doi.org/10.1021/jp970984n>
  
- [29] H. S. Park, S.-H. Park, H. Lee, H.-S. Kim, [Deep learning applied to peak fitting of spectroscopic data in frequency domain](#), *New Phys.: Sae Mulli* 70 (11) (2020) 920. doi:[10.3938/NPSM.70.920](http://www.npsm-kps.org/journal/view.html?doi=10.3938/NPSM.70.920).  
URL <http://www.npsm-kps.org/journal/view.html?doi=10.3938/NPSM.70.920>