

# Self-Supervised Approaches to the Classification of Spectra: Application to Phase Transitions in X-ray Diffraction Data

Yue Sun (✉ [yue.sun@xfel.eu](mailto:yue.sun@xfel.eu))

University of Szeged

Sandor Brockhauser

University of Szeged

Péter Hegedűs

University of Szeged

Christian Plückthun

European XFEL GmbH

Luca Gelisio

European XFEL GmbH

Danilo Enoque Ferreira de Lima

European XFEL GmbH



---

## Article

### Keywords:

**Posted Date:** February 28th, 2023

**DOI:** <https://doi.org/10.21203/rs.3.rs-2599173/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

**Version of Record:** A version of this preprint was published at Scientific Reports on June 9th, 2023. See the published version at <https://doi.org/10.1038/s41598-023-36456-y>.

# Abstract

The ability to detect interesting events is instrumental to effectively steer experiments and maximize their scientific efficiency. To address this, here we introduce and validate three frameworks based on self-supervised learning which are capable of classifying 1D spectral data using a limited amount of labeled data. In particular, in this work we focus on the identification of phase transitions in samples investigated by x-ray diffraction. We demonstrate that the three frameworks, based either on relational reasoning, contrastive learning, or a combination of the two, are capable of accurately identifying phase transitions. Furthermore, we discuss in detail the selection of data augmentations, crucial to ensure that scientifically meaningful information is retained.

## Introduction

Experimental techniques such as spectroscopy and x-ray diffraction are instrumental in investigating matter (see, e.g., Ref.<sup>1-4</sup>). When experiments are performed at modern x-ray facilities, such as x-ray free electron lasers (XFELs), a vast amount of data are potentially collected over short periods of time. For example, at the European XFEL<sup>5</sup> up to 27,000 pulses can be generated in a second. Of these, up to 5,120 can be stored as detector images<sup>6,7</sup>. The ability to rapidly and accurately assess the status of an experiment is essential to maximize its efficiency. As an example, one may want to rapidly identify structural variations as a function of external variables. On the other hand, when analyzing data already collected – potentially up to hundreds of thousands of data sets – it is crucial to be able to employ some automated or semi-automated method capable of extracting interesting features in the data to minimize the usage of experts' time and to maximize the scientific output.

Methods based on machine learning (ML) are ideal for automation of repetitive tasks and identification of patterns in data sets, and several applications to data collected at x-ray facilities have been recently published (see, e.g., Ref.<sup>8,9,10</sup>). When considering 1D spectral data, numerous classification approaches have been developed, including unsupervised clustering method such as spectral clustering<sup>11</sup>, K-Means<sup>12</sup>, Agglomerative clustering<sup>13</sup>, DBSCAN<sup>14</sup>, and supervised ML methods such as k-nearest neighbors<sup>15</sup>, partial least squares discriminant analysis<sup>16,17</sup>, decision trees<sup>18</sup>, random forests<sup>19</sup>, and extreme learning machines<sup>20,21</sup>. However, for traditional unsupervised clustering algorithms, fine-tuning of hyperparameters is usually required to obtain accurate results. This task is typically performed by experts and might need to be reiterated for different data sets, which hinder the automation of data analysis and greatly reduce efficiency. An alternative to increase accuracy and reduce computation time is to employ supervised ML methods. These may require a long training time, and rely on experts' knowledge in terms of data annotation. Current popular methods are based on deep neural networks (DL), of which the most commonly established are convolutional neural network (CNN)<sup>22-24</sup>, recurrent neural networks (RNNs)<sup>25,26</sup>, attention-based neural networks<sup>27,28</sup>, and hybrid models<sup>26,27,29</sup>. However, it should be noted that the strength of supervised ML methods, that is the possibility of introducing domain-knowledge through annotation, is often problem-specific and time-consuming, which again hinders automation. Recently, methods based on self-supervised learning have opened up a new research frontier<sup>30</sup>. These are based on data augmentation and appropriate pretext tasks, through which deep neural networks can learn generalizable features from unlabeled data. Self-supervised methods impose invariance to appropriate

pretext tasks, which are solved with the aim of learning a good representation of the data and designed according to specific problems<sup>31-34</sup>. While self-supervised learning requires domain-specific knowledge, the need for human supervision is largely reduced with respect to supervised learning and the potential for automation is increased. In this study, we focus on two branches of self-supervised learning, that is self-supervised relational reasoning learning<sup>35, 36, 37, 38</sup> and self-supervised contrastive learning<sup>31, 32, 34</sup>.

The relational reasoning networks are based on a key design principle, that is the use of a relation network (usually a multi-layer perception (MLP)) as a learnable function to quantify the relations between entities and their properties<sup>36</sup>. While the relational reasoning paradigm has gained traction in the deep learning community only recently<sup>36</sup>, it has achieved promising results in many fields, for example, video processing<sup>37</sup>, few-shot nature image recognition<sup>38</sup>, and time series data classification<sup>35</sup>. However, its application in the natural sciences is still scarce. Contrastive learning<sup>39</sup> is based on learning similar/dissimilar representations from unlabeled data. The key principle is to extract underlying patterns in data by maximizing similarities of augmentations from the same instances while minimizing the similarity of different instances<sup>31</sup>. Recently, contrastive learning has attracted increasing attention in the natural sciences and has shown remarkable results on a variety of scientific problems, including molecular representation<sup>40,41</sup>, density-of-states of 2D photonic crystals<sup>42</sup>, similarity search for sky surveys<sup>43</sup>, single-particle diffraction images<sup>44</sup>, and Raman spectrum. In particular, Ref.<sup>42</sup> demonstrates that self-supervised contrast learning can greatly reduce the number of labels required to train a network, which is tedious and time-consuming operation. These successful applications in different scientific fields demonstrate the effectiveness and versatility of contrastive learning.

In this work, we demonstrate that self-supervised machine learning methods can provide great opportunities to improve the scientific efficiency of experiments at large-scale x-ray facilities. We explore the application of self-supervised relational reasoning and contrastive learning to 1D spectral classification problems. In particular, we demonstrate that it can be effectively used to classify phase transitions observed in X-ray diffraction (XRD) experiments<sup>45-47</sup>. We introduce and discuss three self-supervised representation learning frameworks for the classification of data, namely SpecRR-Net, SpecMoco-Net, and SpecRRMoco-Net. SpecRR-Net extracts discriminative features from unlabeled spectra based on relational reasoning, which attempts to discover data representations by reasoning the relation among entities<sup>35,36</sup> in multiple dimensions and at different scales. SpecMoco-Net is based on contrastive learning, which aims to build representations by learning similarities and dissimilarities between different objects<sup>31,32</sup>. SpecRRMoco-Net benefits from both relational reasoning and contrastive learning, unifying SpecRR-Net and SpecMoco-Net. The backbone encoders applied in all three models were adapted from the ConvSC attention model in Ref.<sup>27</sup>, which was specifically designed for 1D spectral classification. We furthermore demonstrate the validity and performances of these three frameworks targeting the identification of a phase transition as seen by x-ray diffraction. The results show that SpecRR-Net and SpecRRMoco-Net are superior, and can effectively reduce the time spend by scientists annotating data manually, therefore offering great potential to automate the classification process.

## Methods

In this section, after introducing the case study, we present the proposed self-supervised spectral classification framework, shown in Fig. 3. It includes self-supervised training to learn useful representations from unlabeled

spectral data, and downstream supervised classification based on small amounts of labeled data. Self-supervised learning methods generally include two aspects: pretext tasks and loss functions. A crucial step for the success of self-supervised learning is the definition of proper objectives for unlabeled data in conjunction with data augmentation. In this work, we define four pretext tasks by exploring the meaningful information of 1D spectral data itself. Based on this, four surrogate-objective functions are proposed. In this way, useful representations can be learned by solving these pretext tasks, with the aim of significantly reducing the amount of labels and increasing the automation of the classification process. In the following, we first introduce the formulation of the problem, then detail the data augmentation applied in this work, and finally detail our approach.

**Experimental data.** We validate the proposed self-supervised approaches using data from high-pressure X-ray diffraction experiments. These allow us to probe the evolution of parameters of the atomic lattice and to identify phase transitions, if any, during compression and decompression of the sample. Modifications of the lattice are reflected by changes of the scattering curves (examples of these are shown in Fig. 1), which are the inputs to our framework. In this investigation, we particularly aim at detecting phase transitions, whose fingerprint is typically the appearance of new peaks in the scattering curves. In the rest of the manuscript, XRD data collected from wüstite (FeO) and Fe powder samples at the P02.2 beamline of the synchrotron light source PETRA III<sup>48</sup> at DESY will be used. Details on the experiment are provided in the supplementary material SM-1. We divide measured spectra into three different classes depending on the atomic arrangement within the sample, namely before, during and after the phase transition.

**Problem Definition.** Given unlabeled data containing a series of spectral curves  $\{x_i\}$ , we aim to learn a parametrized map  $f_\theta(\cdot)$ , which can produce a rich and descriptive representation  $z_i = f_\theta(x_i)$  from unlabeled spectra for the downstream classification task. In this equation,  $\theta$  are the learnable parameters of the neural networks. The learned representations will be then used for downstream spectral classification tasks while using a minimal number of labels.

**Data augmentation.** Data augmentation, which provides different views of the input data expected to be mapped to similar representation vectors, is critical in defining useful pretext tasks<sup>31</sup> in self-supervised learning. Such augmentations produce varied spectra, possibly with simulated additional experimental complexity or noise, but still plausible and with the same target labels. The objective function therefore ensures that same-label variations of the input spectra must be represented similarly. Such a procedure increases the robustness and generalization capabilities of the model, as variations of the input dataset are also used to train the model.

In this work, we first preprocessed the spectra data by normalizing them to the  $[0, 1]$  range, then we sequentially applied diffraction angle warping (which is adapted from time warping<sup>49</sup> changing its original time dimension to the diffraction angle dimension), and magnitude warping<sup>49</sup> as data augmentations. Magnitude warping is used to simulate reasonable and random variations in the intensities of peaks, while not changing their positions. Diffraction angle warping is used to parallel the variation of peak positions, so to allow the model to focus more on the number of peaks rather than their location. An example of the effect of the augmentations is shown in Fig. 2. It is important to note that both data augmentations are physically meaningful and specific to the case of study. In fact, neither data augmentation results in changes in the number or shape of the peaks, which are the primary indicators for detecting a phase transition.

**Self-supervised first stage training and second stage 1D spectra classification.** The self-supervised classification framework adopts a two-stage training, as shown in Fig. 3. In the first stage, the feature extraction backbone encoder is trained in an unsupervised manner through momentum contrastive learning and relational reasoning-based learning. The objective of this first stage is to learn useful representations, thus reducing the amount of label information needed. After the first stage training, the contrast head and relational reasoning heads are discarded to reduce the correlation between output variables, as it has been suggested in related self-supervised learning research, such as in Ref.<sup>31,32,35</sup>, the backbone parameters are completely transferred to the second part for downstream classification tasks. In the second stage, a single-layer linear classifier is trained using a reduced amount of labeled data, projecting the latent space to physically meaningful spectral phase classes.

The shared feature extraction backbone model  $f_q$  applied in this approach is the Conv SC attention model from Ref.<sup>27</sup>, but without feed-forward network, as shown in Fig. S1 of the supplementary material. It consists of two convolution modules for extracting local features, and two self-attention modules performed across spatial (diffraction angle) and channel (introduced by the convolutional channels) dimensions to build long-range dependencies of spectra. In this way, latent dependencies and useful representations can be well captured. Furthermore, to accept input data with different feature sizes, we apply the 1D adaptive average pooling instead of the 1D global max pooling operation in the second convolution module, as shown in Fig. S1 of the supplementary material (see Ref.<sup>27</sup> for more details on this model).

In this work, four pretext tasks are proposed to supervise the training of the backbone encoder in the first stage. These include three relational reasoning-based pretext tasks, i.e., an inter-sample relational reasoning module, an intra-sample relational reasoning module, an external-variable relational reasoning module, and one pretext task based on instance-level contrastive learning, as shown in Fig. 3. We name the self-supervised classification framework based only on three relational reasoning modules as SpecRR-Net, the network based only on the contrastive learning module as SpecMoco-Net, and the combination of these two networks as SpecRRMoco-Net. We will describe each module in detail in the following sections.

**Inter-Sample Relational Reasoning.** The Inter-Sample relational reasoning module<sup>35,36</sup> learns to quantify the relationships of the sampled pairs (how spectral instances are related to themselves and other instances), by formulating it as a binary classification pretext task, as shown in the upper branch of Fig. 4. This pretext task is trained by minimizing the binary cross-entropy loss  $L_{Inter}$ , which follows the loss function of Eq. (1) in Ref.<sup>35</sup>.

**Intra-Sample Relational Reasoning.** The Intra-Sample relational reasoning module<sup>35</sup> models the relation between different spectral pieces within each individual diffractogram. It is adapted<sup>35</sup> from the intra-temporal relational reasoning module in Ref.<sup>35</sup>, originally proposed to model the global temporal dependencies of time series data. Here, we extend it to the diffraction angle dimension. We formulate the intra-sample relational reasoning module as a multi-class classification task trained with cross-entropy loss  $L_{Intra}$ , which follows the loss function of Eq. (2) in Ref.<sup>35</sup>. The hyperparameters are consistent with those of the module in Ref.<sup>35</sup>. In this way, the underlying dependencies along the diffraction angle dimension can be captured. Further details are given in the SM-2 of the supplementary material.

**External Variable Relational Reasoning.** Although the above two relational reasoning modules can learn latent discriminative features from sampled pairs, they do not properly utilize information on external variable of spectra. The external variables could be any external condition in the experiment, e.g., temperature, pressure, or electric field. In our specific use case, this is pressure, which varies with time during compression and decompression. Guided by this, we designed and introduced a third relational reasoning branch (see Fig. 3 and Fig. 4), that is the external variable relational reasoning module, to build robust external dependencies from the spectral samples. This can further enable the backbone to learn useful patterns along the external variable dimension.

Formally, given any spectral curve  $x_i$  collected at time step  $i$ , the encoded representation of its augmented version  $x_i^A$  is denoted by  $z_i = f_\theta(x_i^A)$ . A single layer external-variable relation projection head  $r_\gamma(\cdot)$  is applied to reason the external variable relation score, denoted as  $s_i^{tmp} = r_\gamma(z_i)$ . First, the spectral curves are evenly divided into  $T=5$  external variable relation categories in order of acquisition time. Then, a multi-class classification pre-task is constructed and trained with the cross-entropy loss  $L_{tmp}$  as

$$L_{tmp} = -\frac{1}{B} \sum_{i=0}^{B-1} \log \frac{\exp(s_{i,y_i}^{tmp})}{\sum_{t=0}^{T-1} \exp(s_{i,t}^{tmp})}, \quad (1),$$

where  $y_i^{tmp}$  is the ground-truth class label in this module. The ablation study of hyperparameters  $T$  is left to future work.

The relational scores in these three relational reasoning modules are not fixed distance metric but rather a learnable nonlinear similarity measure for comparing the similarity of encoded spectra data.

**Self-Supervised contrastive learning module for 1D spectra classification.** In the self-supervised contrastive learning module, instance-wise contrastive learning<sup>31,34</sup> is employed, where each spectra instance is treated as a distinct class of its own and a pretext classifier is trained to distinguish between individual instances<sup>50</sup>. SpecMoco-Net is based on momentum contrastive learning (MoCo)<sup>34</sup>. This is formulated as minimizing InfoNCE-based contrastive loss function  $L_{cont}$  proposed by Ref.<sup>31,34</sup>. During training, the unsupervised contrastive loss brings spectra containing similar spectral peak features closer together in latent space, while spectra with different spectral features are pushed farther apart. Dissimilarities, within our case of study, are, e.g., different number of peaks, at different positions, or with different shapes. Further details are given in the SM-2 of the supplementary material.

**Self-Supervised loss function.** As can be seen from Fig. 3, the above four modules share the same backbone encoder  $f_q$ . The training of the shared feature extraction encoder can also be viewed as multi-task learning. By jointly optimizing the inter-sample relational reasoning objective, the intra-sample relation reasoning objective, the external-variable relational reasoning objective, and the self-contrastive learning objective, the final training loss function of SpecRRMoco-Net is specified as

$$L = L_{inter} + L_{intra} + L_{tmp} + c \cdot L_{cont}. \quad (2),$$

Here  $c$  is a coefficient to adjust the weight of the contrastive loss and is set to 0.01. Ablation studies on this coefficient are presented in 'Experiments and Results' section. It is important to note here that this loss function unifies SpecRR-Net and SpecMoco-Net, i.e., a value of 0 for the coefficient  $c$  corresponds to SpecRR-Net, while a value of infinity (achieved by retaining the  $L_{cont}$  item with a coefficient of 1, while excluding the other three relational reasoning-based loss items in the loss function) is equivalent to SpecMoco-Net).

## Experiments And Results

**Implementation details.** The SpecRRMoco model, SpecRR-Net, and SpecMoco-Net were trained using PyTorch on a single NVIDIA A100-PCIE-40GB. The self-supervised backbone encoder  $f_q(\cdot)$  is trained by minimizing the proposed joint loss function Eq. (2) with a stochastic gradient descent (SGD) optimizer<sup>51</sup>. We named the nine data sets used as D1 to D9 for convenience, where D8 and D9 were collected from FeO powder, and D1-D7, from Fe powder sample. Among them, D1, D4, D8, and D9 data sets (872 spectral curves in total), were used to train the encoder network  $f_q(\cdot)$  to learn feature representations during the first stage training (without label information). The batch size was set to 512, and the capacity of the queue of keys in the contrastive learning module to  $872 \times 2$  (that is, twice the total number of spectral curves in the first stage training data sets). Within the SpecRRMoco-Net framework, experiments were performed with a loss factor of  $c = 0.01$ , unless otherwise stated. We applied data augmentations randomly 6 times in the inter-sample relational reasoning branch. In the first stage, the initial learning rate of the optimizer was set to 0.15, a linear warmup for the first 50 epochs (from a value of 0.02) followed by a cosine decay schedule was applied to adjust the learning rate during training, and the weight decay was set to  $1 \times 10^{-4}$ . While the relation scores in relational reasoning modules are similarity-based, we formulate each relational reasoning pretext task as a classification task, so accuracy-based metrics can be applied to evaluate its performance. See Fig S4 for more details.

In the second stage, a linear was trained by minimizing the cross-entropy loss function, and an SGD optimizer with a learning rate of 0.15, and weight decay of  $1 \times 10^{-4}$  was applied. To further prevent overfitting, a train/validation/test strategy and an early stopping strategy, which stops the training when the validation accuracy does not increase relatively to its previous best value for  $M = 20$  steps, were employed to train the linear classifier. Some representative scattering curves of each phase of our samples were selected as the basis of the training/validation/test dataset, and a total of 42 original curves (2.7% of total) were labeled. Further details are provided in Supplementary Material SM-3.

**Linear evaluation on downstream classification task.** In this subsection, we evaluated the performance of the self-supervised encoder trained by different networks on the downstream spectral classification task. To do so, we train a linear classifier on top of learned representations from the backbone encoder. The representative labeled spectra from five data sets (the four data sets used in the first stage training plus D5) were used to train the linear classifier. Figure 5 shows the classification results together with the classification probabilities of SpecRRMoco-Net for 4 data sets (D2, D3, D8 and D9) using only 2.7% of the labeled data. The classification results of the other 5 data sets are reported in See Fig. S6 of the supplementary material. To simplify interpretation, we only present results around the phase transition zone. The green line represents class 0 (before phase transition), the red line represents class 1 (during phase transition), and the blue line represents class 2 (after phase transition). We compared the classification accuracy of SpecRRMoco-Net (average accuracy 98.6%), SpecRR-Net (average accuracy 98.9%), and SpecMoco-Net (average accuracy 97.9%), as

reported for each dataset in Table 1. The prediction time for each spectral curve is about 40  $\mu$ s, which is small enough to meet the requirement of real-time processing even at high-repetition rate facilities like the European XFEL. The overall results show that with only 2.7% of the labels (42 spectral curves), all three models were able to accurately detect phase transitions. In particular, SpecRR-Net achieved a slightly better classification performance than SpecRRMoco-Net under the current training strategy and hyperparameter settings, while SpecMoco-Net performed worse among the three models. In SpecMoco-Net, the predicted class labels of some spectra in D8 data set were inconsistent around the 'during phase transition' region (Fig. S7). Also, a few spectral curves were misclassified in a few other data sets. It should be noted that the other four data sets (D2, D3, D6, and D7 data sets) were not included in the self-supervised learning or the linear evaluation, but nevertheless the self-supervised models still achieved very good classification results, meaning that the learned representation is transferable. Moreover, it also demonstrates the high quality of the learned representations of the feature extraction backbone encoder.

**Comparison with other methods.** In this section, we compare the self-supervised classification models already introduced with a modified version of the SelfTime network<sup>35</sup> (designed specifically for time series data), which we name SpecSelfTime. In particular, we replaced the original convolutional backbone encoder with the ConvSC attention network to better fit 1D spectral data for better performance. It should also be noted that SpecSelfTime, which is closely related to our work and is the baseline of our SpecRR-Net, does not include the external variable relational reasoning module we introduced in this study.

For a fair comparison, the settings of hyperparameters in the SpecSelfTime model are the same as in the SpecRR-Net and SpecRRMoco-Net models. Table 1 show its classification results on experimental spectra with 2.7% labeled data (average accuracy 92.6%). SpecSelfTime performs poorly on several data sets, and particularly on D8 and D3, where it failed to detect the 'during phase transition' class on D8 and failed to distinguish the 'before-' and 'after-phase transition' categories on D3. This indicates poor generalization ability of the model. More classification results are presented in Fig. S7 of the supplementary material. As can be seen from the results, SpecSelfTime performs worse than the improved SpecRR-Net and even SpecMoco-Net, which highlights the importance of the external-variable relational reasoning module we introduced.

We furthermore considered the unsupervised spectral clustering method Spectral Clustering<sup>11</sup>. Also, in this case, classification results for each dataset are shown in Table 1. As it can be observed, it performs well on data sets with abrupt phase transitions (D1-D7, although slightly worse than the three self-supervised methods we have introduced), but not so well on the 'during phase transition' regions of the D8 and D9 datasets.



Table 1

Classification accuracy with different methods. For all self-supervised methods, 2.7% of labeled data were used.

Model	D1	D2	D3	D4	D5	D6	D7	D8	D9	Average
Spectral Clustering	0.950	0.983	0.971	0.969	0.985	0.971	0.992	0.734	0.670	0.914
SpecRRMoco + Spectral Clustering	0.983	0.991	0.986	<b>0.992</b>	<b>0.992</b>	<b>1.000</b>	<b>1.000</b>	0.775	0.896	0.957
SpecSelfTime	<b>1.000</b>	<b>1.000</b>	0.700	<b>0.992</b>	0.900	0.986	<b>1.000</b>	0.793	0.959	0.926
SpecRR-Net	0.983	0.991	0.986	<b>0.992</b>	<b>0.985</b>	<b>1.000</b>	<b>1.000</b>	<b>0.986</b>	<b>0.980</b>	<b>0.989</b>
SpecMoco	0.983	0.948	<b>1.000</b>	0.985	0.992	<b>1.000</b>	<b>1.000</b>	0.955	0.950	0.979
SpecRRMoco	<b>1.000</b>	0.991	0.986	0.985	0.977	<b>1.000</b>	<b>1.000</b>	0.973	0.961	<b>0.986</b>

While the downstream classification task can evaluate the quality of the model, it cannot fully reflect the clustering ability. Therefore, as a qualitative analysis, we further evaluate the clustering power of these self-supervised classification models by visualizing the learned representations using t-SNE (t-Distributed Stochastic Neighbor Embedding)<sup>52</sup>. Figure 6a renders the t-SNE of SpecRRMoco-Net, while Fig. 6b visualizes the original data with labels predicted by the spectral clustering method. The t-SNE visualization plots of SpecRR-Net, SpecMoco-Net, and SpecSelfTime are given in Fig. S8 of the supplementary material. The corresponding silhouette coefficients of each self-supervised model are shown in Table 2. These are calculated after predicting labels of each data sets by applying unsupervised spectral clustering to the learned representations of each self-supervised encoder. The average silhouette values were 0.798 for SpecRRMoco-Net, 0.791 for SpecSelfTime, 0.780 for SpecMoco-Net, and 0.821 for SpecRR-Net. Compared to SpecRRMoco network, SpecRR-Net has the better clustering ability, effectively amplifying the differences between different classes while clustering data from the same class together. SpecRRMoco-Net and SpecRR-Net exhibit better clustering ability than SpecMoco-Net, which is consistent with the classification results.

To verify whether the self-supervised model works well when no label information is available, we applied the spectral clustering method to the learned representation of the SpecRRMoco encoder (as it unifies SpecRR-Net and SpecMoco-Net). The classification results for each dataset are listed in Table 1, and the corresponding silhouette values in Table 2. It should be noted that in this experiment, class labels are assigned by the spectral clustering method without using data annotations, so it can be regarded as a completely unsupervised learning method. We found that the proposed combination of SpecRRMoco-Net and spectral clustering performs better than Spectral Clustering alone, especially on the data sets with ‘during phase transition’ regions (D8 dataset and D9 dataset). The classification accuracy and Silhouette Coefficient values show that our proposed method can extract useful underlying patterns from unlabeled data, thus making the following classification tasks easier. This can significantly overcome the label sparsity problem and greatly automate the spectral classification procedure.

Table 2  
Average Silhouette coefficient obtained by applying different methods to each data set.

Model	D1	D2	D3	D4	D5	D6	D7	D8	D9	Average
Spectral Clustering (SC)	0.684	0.623	0.800	0.793	0.616	0.729	0.923	0.594	0.576	0.704
SpecSelfTime + SC	<b>0.825</b>	<b>0.856</b>	0.923	<b>0.924</b>	0.845	<b>0.861</b>	0.955	0.422	0.506	0.791
SpecRRMoco (c 0.01) + SC	0.817	0.840	0.916	0.879	0.784	0.854	0.961	0.537	<b>0.593</b>	0.798
SpecRR-Net + SC	<b>0.824</b>	0.831	<b>0.926</b>	0.866	<b>0.849</b>	<b>0.858</b>	<b>0.971</b>	<b>0.674</b>	0.588	<b>0.821</b>
SpecMoco + SC	0.798	0.722	0.895	0.882	0.813	0.741	0.920	0.658	<b>0.592</b>	0.780

**Ablation studies on the coefficient in the SpecRRMoco-Net loss function.** Here, we report on an ablation study on the coefficient (shown in Table 3), performed to understand its impact on learning data representations. These experiments were performed under the same training setup described above. We varied in the range  $[0.001, 1]$ , and also set it to 0 (that is, a pure SpecRR-Net) and infinity (that is, a pure SpecMoco-Net). For the downstream spectral classification task, 2.7% of labels were used. From Table 3, we can see that SpecRRMoco-Net performs well over a wide range of the coefficient (0-0.1 and infinity). This result suggests that jointly optimizing the relational reasoning-based pretext task and the contrast learning-based pretext task can improve the performance of the pure contrast learning-based network, but not of the purely self-supervised relational reasoning network under the current training setup.

Table 3  
Ablation study of the coefficient  $c$  in the loss function.

c	D1	D2	D3	D4	D5	D6	D7	D8	D9	Average
0.001	0.983	0.931	0.957	0.992	1.000	0.971	0.988	0.968	0.911	0.967
0.01	1.000	0.991	0.986	0.985	0.977	1.000	1.000	0.973	0.961	<b>0.986</b>
0.1	0.983	0.983	1.000	0.985	1.000	1.000	1.000	0.950	0.943	0.983
1	0.967	0.914	0.943	0.985	0.938	0.986	1.000	0.919	0.965	0.957
0 (SpecRR-Net)	0.983	0.991	0.986	0.992	0.985	1.000	1.000	0.986	0.980	<b>0.989</b>
Inf (SpecMoco-Net)	0.983	0.948	1.000	0.985	0.992	1.000	1.000	0.955	0.950	0.979

**Ablation studies on the data augmentation.** We report here on an ablation study on data augmentations performed in order to evaluate their impact on the SpecRRMoco-Net performances. Several commonly used

data augmentation techniques were explored, including diffraction angle warping, magnitude warping, window slicing, jitter, and scaling. Among them, diffraction angle and window slicing are performed in the diffraction angle dimension, whereas jitter, scaling, and magnitude warping are performed in the magnitude domain. These data augmentation techniques introduce invariances with respect to the physics knowledge. They generate new input with variances while keeping identical labels in the embedding space. Based on this, surrogate tasks can be formed to extract underlying patterns and build the representations. In addition to the diffraction angle and magnitude warping which were already discussed, jittering was used to introduce possible random noise in the experiment, scaling was used to model uniform intensity variations, and window slicing was used to model small variations in diffraction angle coverage. As can be seen from Fig. 7, diffraction angle warping is the most important data augmentation technique in this application, since the number of diffraction peaks is crucial in this experiment. In addition, when this is combined with the transformation in the magnitude domain, performances are improved. In particular, the best result is achieved when diffraction angle is combined to magnitude warping. Therefore, in this study, we applied amplitude warping and diffraction angle sequentially to all models presented in this paper. This experiment illustrates that data augmentations play an important role in self-supervised models<sup>31,32</sup>. As it is domain-specific, it must be customized for data sets from different research areas. Once the most appropriate data augmentation techniques are identified, the ability to automatically classify the data can be effectively improved.

## Discussion

From the above experimental results, it can be concluded that the three networks proposed in this paper are effective in constructing data representations that can greatly improve the automation of the classifications of spectral data, and in particular the detection of phase transitions. We attribute the success of the models, consistent with the results of the ablation study, to appropriate data augmentations and pretext tasks. In fact, self-supervised learning critically relies on augmentations, which should be tailored for the scientific case object of investigation. The ones applied in this study retain physically meaningful information while simulating other plausible experimental effects. Thus, compared to traditional unsupervised clustering algorithms which require manual tuning of parameters for each dataset, self-supervised models allow the classification process to be automated once a minimal amount of labels is available.

In SpecMoco-Net, the learning process is primarily based on exploiting biases in the data, rather than learning to perform inference tasks based on the data itself. In addition, SpecMoco-Net is based on the instance-instance discrimination task, which cannot explicitly exploit data information at different scales, such as the global dependencies across diffraction angle dimension. Furthermore, in practice, self-supervised contrastive learning benefits from a large number of negative samples to extract meaningful representations, and while SpecMoco-Net allows a large and consistent dynamic dictionary, in our case we do not have enough spectral training examples, which is another important reason why SpecMoco-Net performed worse than SpecRR-Net and SpecRRMoco-Net in our case of study.

Networks based on relational reasoning learning can be viewed as simultaneously learning deep embeddings and deep non-linear metrics (similarity functions)<sup>38</sup>. In SpecRR-Net and SpecRRMoco-Net, three relational reasoning modules are designed to capture the underlying dependencies from multiple dimensions and at different scales to build useful representations. Moreover, comparison with SpecSelfTime shows that our

proposed external-variable relational reasoning module can significantly improve the performance of models by addressing the dependencies of diffraction spectra on pressure values, in this particular application. Relative to the pretext task based on contrast learning, the relational reasoning-based pretext tasks impose more supervision on the network using easily accessible sources of information. In the process of reasoning about the relations between spectral entities, irrelevant and noisy features are neglected, and non-obvious properties can be focused on, thereby gaining new knowledge. Furthermore, the difference in the structure of the two methods may also lead to some differences in the way of updating model parameters. Ablation studies on structural differences are necessary and interesting for further research, which is left to future work.

SpecRRMoco-Net benefits from both relational reasoning learning and contrastive learning, and although it does not show better results than SpecRR-Net with the current hyperparameters and training settings, it combines SpecRR-Net and SpecMoco-Net therefore providing a flexible framework that can potentially fit a broader set of use cases. The success of each pre-text task in SpecRRMoco-Net drives the update of the encoder model, improving its feature representation ability while increasing the robustness and generality of the encoder network. Importantly, although these models are proposed for classification applications on spectral data, the architectures are general and can be easily extended to 1D time series data and various other types of data, such as image classification.

Further evaluation and interpretation of the model are given in SM-5 of the supplementary material.

## Conclusions

In this paper we propose three self-supervised frameworks to classify 1D spectral data using a minimal amount of labeled data, and we validate their performance using x-ray diffraction data of samples showing phase transitions. These frameworks are based on relational reasoning (SpecRR-Net), contrastive learning (SpecMoco-Net) or a linear combination of the two (SpecRRMoco-Net). They are capable of learning discriminative features and building effective representations, therefore greatly reducing the number of labels required, making a step towards automating the spectral classification process. Moreover, as a consequence of the reduced number of labels, scientist's time is greatly optimized. In order to account for the relation between spectra collected along some external variable, we extend the relational reasoning-based method to explicitly include it. In this work, we demonstrate the importance of a proper choice of data augmentations, which must be tailored for the specific case of study to ensure the retention of scientifically meaningful information. In particular, we discuss and validate augmentations relevant to the case study discussed, and we prove that the three methods introduced are effective in detecting phase transitions. This is the case even when data for which no labels are available are used, which demonstrates good generalizability of the approaches. We furthermore compare the three frameworks with state-of-the-art unsupervised methods.

After an initial training step, the methods proposed here can be used to accurately and automatically screen collected data, even in real-time at high-repetition rate facilities given the inference speed, so to provide a better understanding of the experiment and therefore enable the most effective real-time planning.

In future research, we will further validate our self-supervised classification models on spectral data collected from different experiments and spectroscopy techniques. In addition, we plan to work on an automated way of optimizing hyperparameter settings, training strategies and augmentations.

# Declarations

## Acknowledgments

This work was supported by the China Scholarship Council (CSC, No.201904890020), the European XFEL, and the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory. Furthermore, Sandor Brockhauser was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project 460197019 and Christian Plückthun was funded by DFG project No: KO-5262/1. We acknowledge DESY (Hamburg, Germany), a member of the Helmholtz Association HGF, for the provision of experimental facilities. Parts of this research were carried out in P02.2 Extreme Conditions Beamline (ECB) and we would like to thank the beamline staff for their assistance during the experiments. Beamtime was allocated for proposals I-20190025 and I-20190818. D.E.F.D.L and L.G. acknowledge the internal R&D funds supporting machine learning activities at the EuXFEL. The authors acknowledge James Wrigley for his careful reading and checking of the manuscript.

## Author contributions

Y.S.: Study conceptualization, methodology development, software development, data analysis, manuscript writing. S.B.: Study conceptualization, data management review, manuscript writing. P.H.: Study conceptualization, software engineering review, manuscript writing. C.P.: Study conceptualization, data collection, data interpretation. L.G.: Methodology development, and manuscript writing. D.E.F.D.L: Study conceptualization, methodology development, validation, manuscript writing. All authors reviewed the manuscript and agreed to its published version.

## Data availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Additional information

Correspondence and requests for materials should be addressed to Y.S., P.H, and D.E.F.D.L.

## Competing interests

The authors declare no competing interests.

# References

1. Zimmermann, P. *et al.* Modern X-ray spectroscopy: XAS and XES in the laboratory. *Coord. Chem. Rev.***423**, 213466 (2020).
2. Shen, G. & Mao, H. K. High-pressure studies with x-rays using diamond anvil cells. *Rep. Prog. Phys. Phys. Soc. G. B.***80**, 016101 (2017).
3. Peterson, V. K., Auckett, J. E. & Pang, W.-K. Real-time powder diffraction studies of energy materials under non-equilibrium conditions. *IUCrJ***4**, 540–554 (2017).

4. Renner, O. & Rosmej, F. B. Challenges of x-ray spectroscopy in investigations of matter under extreme conditions. *Matter Radiat. Extrem.***4**, 024201 (2019).
5. Decking, W. *et al.* A MHz-repetition-rate hard X-ray free-electron laser driven by a superconducting linear accelerator. *Nat. Photonics***14**, 391–397 (2020).
6. Allahgholi, A. *et al.* AGIPD, a high dynamic range fast detector for the European XFEL. *J. Instrum.***10**, C01023 (2015).
7. Veale, M. C. *et al.* Characterisation of the high dynamic range Large Pixel Detector (LPD) and its use at X-ray free electron laser sources. *J. Instrum.***12**, P12003 (2017).
8. Zhuang, Y. *et al.* Unsupervised learning approaches to characterizing heterogeneous samples using X-ray single-particle imaging. *IUCr***9**, 204–214 (2022).
9. Ignatenko, A. *et al.* Classification of diffraction patterns in single particle imaging experiments performed at x-ray free-electron lasers using a convolutional neural network. *Mach. Learn. Sci. Technol.***2**, 025014 (2021).
10. Assalauova, D., Ignatenko, A., Isensee, F., Trofimova, D. & Vartanyants, I. A. Classification of diffraction patterns using a convolutional neural network in single-particle-imaging experiments performed at X-ray free-electron lasers. *J. Appl. Crystallogr.***55**, 444–454 (2022).
11. Jia, H., Ding, S., Xu, X. & Nie, R. The latest research progress on spectral clustering. *Neural Comput. Appl.***24**, 1477–1486 (2014).
12. Hartigan, J. A. & Wong, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.***28**, 100–108 (1979).
13. Murtagh, F. & Legendre, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J. Classif.***31**, 274–295 (2014).
14. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. in *Proc. of 2nd International Conference on Knowledge Discovery and* 226–231 (1996).
15. Zhang, S., Li, X., Zong, M., Zhu, X. & Cheng, D. Learning k for kNN Classification. *ACM Trans. Intell. Syst. Technol.***8**, 43:1–43:19 (2017).
16. Chen, H., Lin, Z. & Tan, C. Nondestructive Discrimination of Pharmaceutical Preparations Using Near-Infrared Spectroscopy and Partial Least-Squares Discriminant Analysis. *Anal. Lett.***51**, 564–574 (2018).
17. Song, W., Wang, H., Maguire, P. & Nibouche, O. Nearest clusters based partial least squares discriminant analysis for the classification of spectral data. *Anal. Chim. Acta***1009**, 27–38 (2018).
18. Suzuki, Y. *et al.* Symmetry prediction and knowledge discovery from X-ray diffraction patterns using an interpretable machine learning approach. *Sci. Rep.***10**, 21790 (2020).
19. Menze, B. H. *et al.* A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics***10**, 213 (2009).
20. Zheng, W., Shu, H., Tang, H. & Zhang, H. Spectra data classification with kernel extreme learning machine. *Chemom. Intell. Lab. Syst.***192**, 103815 (2019).
21. Zheng, W., Fu, X. & Ying, Y. Spectroscopy-based food classification with extreme learning machine. *Chemom. Intell. Lab. Syst.***139**, 42–47 (2014).

22. Maffettone, P. M. *et al.* Crystallography companion agent for high-throughput materials discovery. *Nat. Comput. Sci.***1**, 290–297 (2021).
23. Lee, J.-W., Park, W. B., Lee, J. H., Singh, S. P. & Sohn, K.-S. A deep-learning technique for phase identification in multiphase inorganic compounds using synthetic XRD powder patterns. *Nat. Commun.***11**, 86 (2020).
24. Zahid, M. U. *et al.* Robust R-Peak Detection in Low-Quality Holter ECGs Using 1D Convolutional Neural Network. *IEEE Trans. Biomed. Eng.***69**, 119–128 (2022).
25. Dandil, E. & Karaca, S. Detection of pseudo brain tumors via stacked LSTM neural networks using MR spectroscopy signals. *Biocybern. Biomed. Eng.***41**, 173–195 (2021).
26. Wang, P. *et al.* Discrimination of blood species using Raman spectroscopy combined with a recurrent neural network. *OSA Contin.***4**, 672–687 (2021).
27. Sun, Y., Brockhauser, S. & Hegedűs, P. Comparing End-to-End Machine Learning Methods for Spectra Classification. *Appl. Sci.***11**, 11520 (2021).
28. Rußwurm, M. & Körner, M. Self-attention for raw optical Satellite Time Series Classification. *ISPRS J. Photogramm. Remote Sens.***169**, 421–435 (2020).
29. Pomyen, Y. *et al.* Deep metabolome: Applications of deep learning in metabolomics. *Comput. Struct. Biotechnol. J.***18**, 2818–2825 (2020).
30. Ohri, K. & Kumar, M. Review on self-supervised image recognition using deep neural networks. *Knowl.-Based Syst.***224**, 107090 (2021).
31. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. in *Proceedings of the 37th International Conference on Machine Learning* 1597–1607 (PMLR, 2020).
32. Chen, X., Fan, H., Girshick, R. and He, K. Improved baselines with momentum contrastive learning. Preprint at <https://arxiv.org/abs/2003.04297> (2020).
33. Jaiswal, A., Babu, A. R., Zadeh, M. Z., Banerjee, D. & Makedon, F. A Survey on Contrastive Self-Supervised Learning. *Technologies***9**, 2 (2021).
34. He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9729–9738 (2020).
35. Fan, H., Zhang, F. and Gao, Y. Self-supervised time series representation learning by inter-intra relational reasoning. Preprint at <https://arxiv.org/abs/2011.13548> (2020).
36. Patacchiola, M. and Storkey, A.J. Self-supervised relational reasoning for representation learning. *Advances in Neural Information Processing Systems*. **33**, 4003–4014 (2020).
37. Zhou, B., Andonian, A., Oliva, A. & Torralba, A. Temporal Relational Reasoning in Videos. in *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part I* 831–846 (Springer-Verlag, 2018). doi:10.1007/978-3-030-01246-5\_49.
38. Sung, F. *et al.* Learning to Compare: Relation Network for Few-Shot Learning. *2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* 1199–1208 (2018) doi:10.1109/CVPR.2018.00131.
39. Hadsell, R., Chopra, S. & LeCun, Y. Dimensionality Reduction by Learning an Invariant Mapping. *2006 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Vol. 2 CVPR06***2**, 1735–1742 (2006).

40. Wang, Y., Wang, J., Cao, Z. & Barati Farimani, A. Molecular contrastive learning of representations via graph neural networks. *Nat. Mach. Intell.***4**, 279–287 (2022).
41. Ji, Z., Shi, R., Lu, J., Li, F. & Yang, Y. ReLMole: Molecular Representation Learning Based on Two-Level Graph Similarities. *J. Chem. Inf. Model.***62**, 5361–5372 (2022)
42. Loh, C., Christensen, T., Dangovski, R., Kim, S. and Soljačić, M. Surrogate-and invariance-boosted contrastive learning for data-scarce applications in science. *Nat. Commun.***13**, 4223 (2022).
43. Stein, G., Harrington, P., Blaum, J., Medan, T. and Lukic, Z. Self-supervised similarity search for large scientific datasets. Preprint at <https://arxiv.org/abs/2110.13151> (2021).
44. Zimmermann, J., Beguet, F., Guthruf, D., Langbehn, B. and Rupp, D. Finding the semantic similarity in single-particle diffraction images using self-supervised contrastive projection learning. Preprint at <https://arxiv.org/abs/2208.11752> (2022).
45. Zhang, Y.Y. et al. Ultrafast X-Ray Diffraction Visualization of B 1– B 2 Phase Transition in KCl under Shock Compression. *Phys. Rev. Lett.***127**, 045702 (2021).
46. Kirschner, M. S. *et al.* Photoinduced, reversible phase transitions in all-inorganic perovskite nanocrystals. *Nat. Commun.***10**, 504 (2019).
47. Ozawa, H., Takahashi, F., Hirose, K., Ohishi, Y. & Hirao, N. Phase Transition of FeO and Stratification in Earth's Outer Core. *Science***334**, 792–794 (2011).
48. Liermann, H.-P. *et al.* The Extreme Conditions Beamline P02.2 and the Extreme Conditions Science Infrastructure at PETRA III. *J. Synchrotron Radiat.***22**, 908–924 (2015).
49. Um, T.T. *et al.* Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. in *Proceedings of the 19th ACM International Conference on Multimodal Interaction* 216–220 (Association for Computing Machinery, 2017). doi:10.1145/3136755.3136817.
50. Wu, Z., Xiong, Y., Yu, S.X. and Lin, D. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3733–3742 (2018).
51. Bottou, L. Large-Scale Machine Learning with Stochastic Gradient Descent. in *Proceedings of COMPSTAT'2010* (eds. Lechevallier, Y. & Saporta, G.) 177–186 (Physica-Verlag HD, 2010). doi:10.1007/978-3-7908-2604-3\_16.
52. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.***9**, 2579–2605 (2008).

## Figures



Representative spectra of different phases

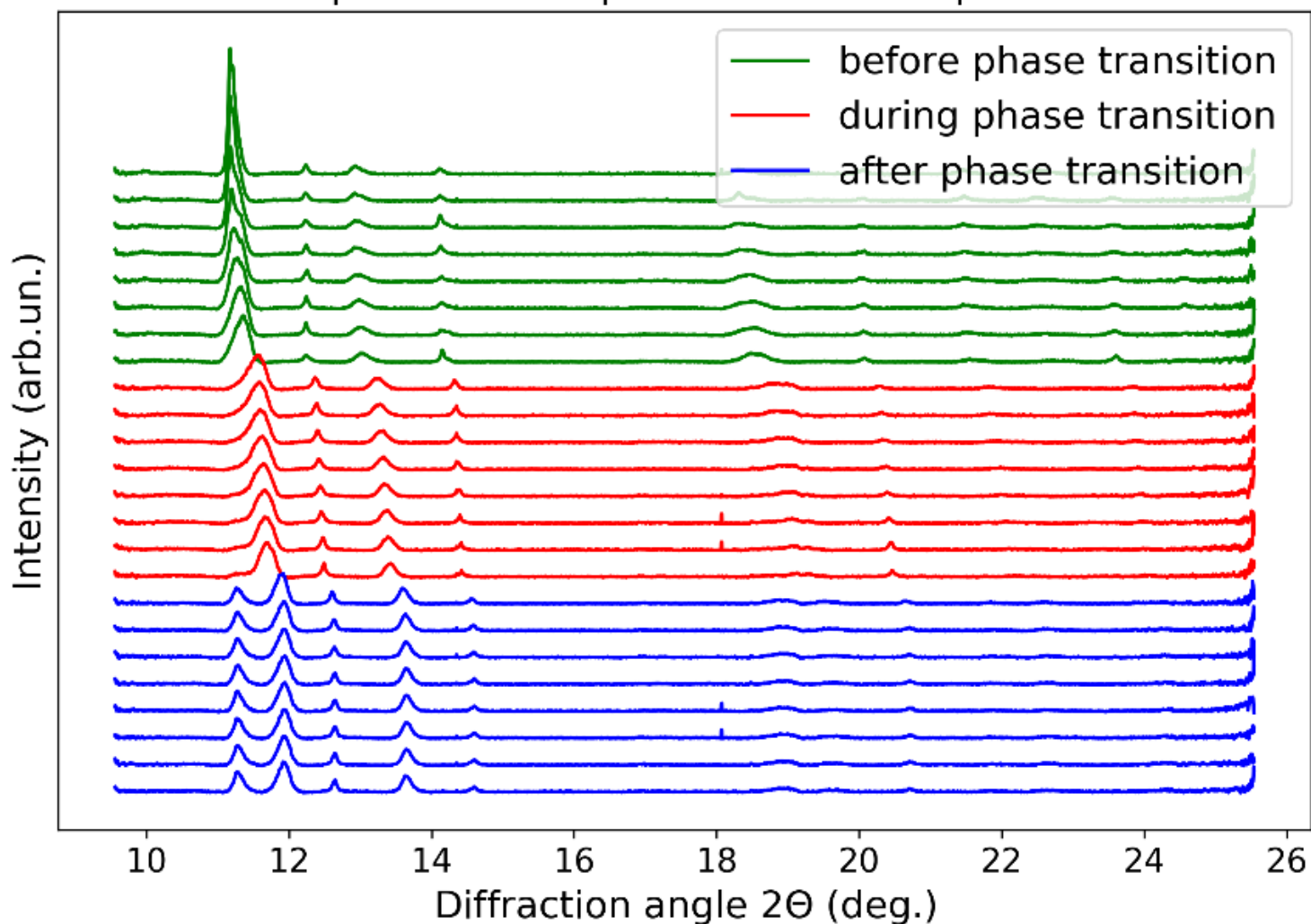


Figure 1

Representative diffraction spectra collected during the experiment on pure FeO powder sample. The spectral curves in green are the ones corresponding to the original phase (before phase transition), the red ones are collected during the phase transition, and the blue ones after the phase transition. Diffraction curves are shifted vertically to improve visualization.

## Effects of applying data augmentation

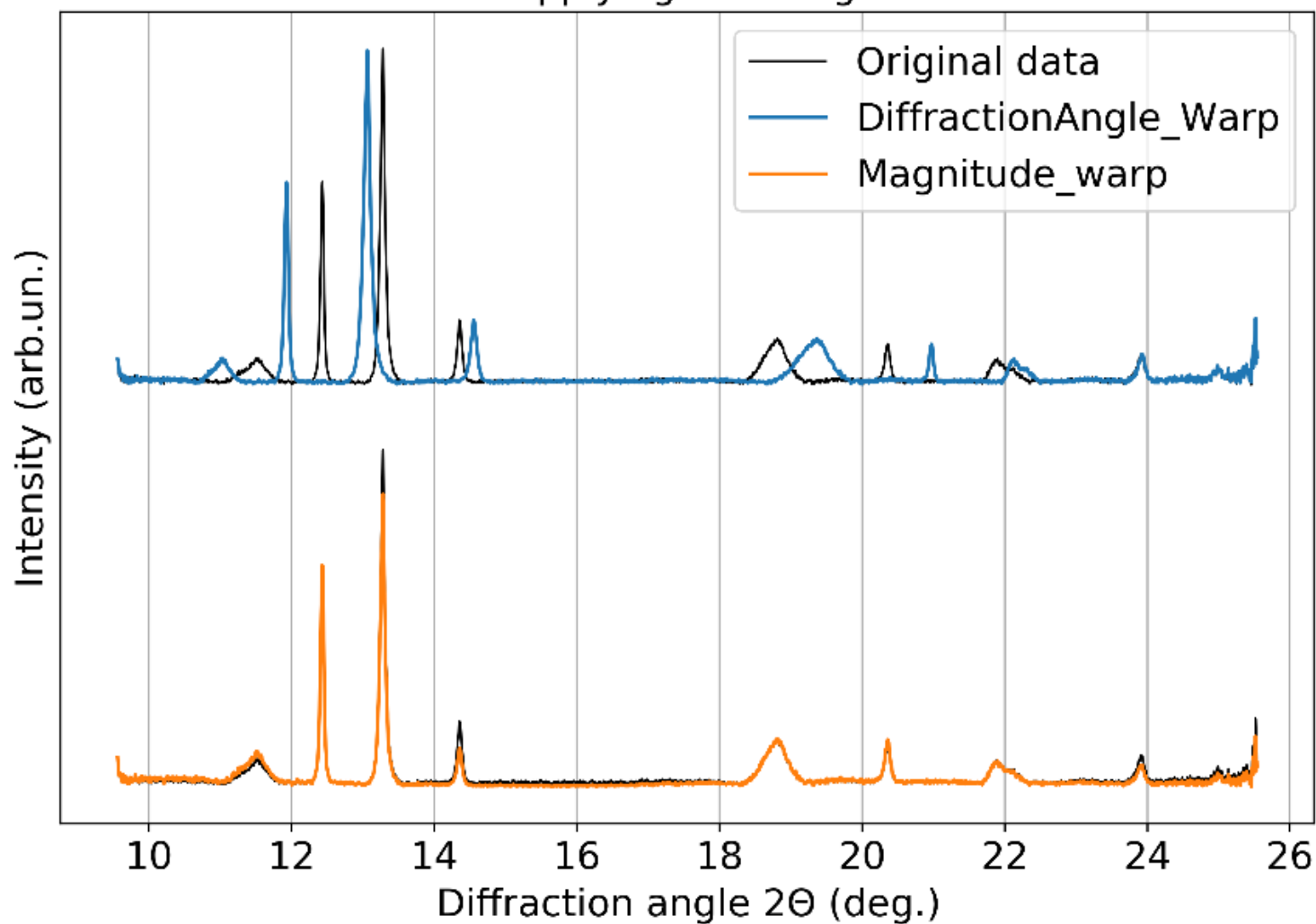


Figure 2

Example of 1D scattering curve, and the effect of applying magnitude warping and diffraction angle warping data augmentations to diffraction spectra.

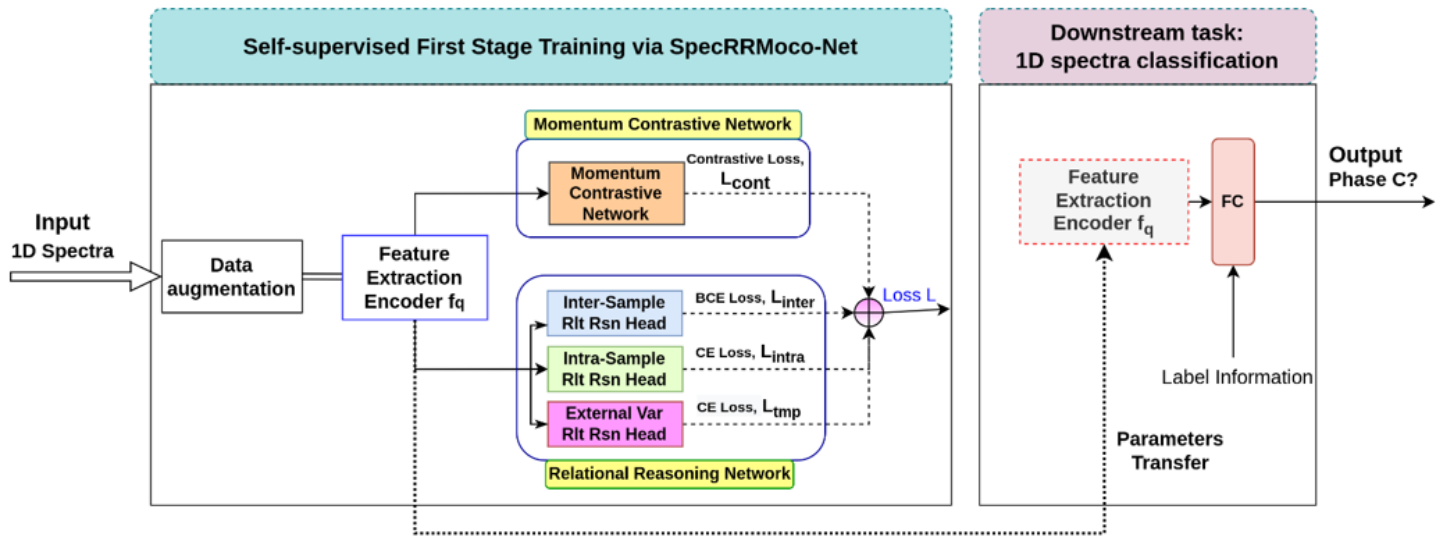


Figure 3. Illustration of the proposed 1D spectra classification framework based on the self-supervised SpecRRMoco-Net, which is a combination of Relational Reasoning Network (SpecRR-Net) and Momentum Contrast Network (SpecMoco-Net). The classification framework consists of two parts, i.e., in the first stage, the encoder  $f_q$  is trained on unlabeled data to build useful representations, and in the second stage, a small number of labels are used to perform the downstream spectral classification task. Specifically, the encoder is training by jointly minimizing the contrastive loss  $L_{cont}$  in SpecMoco-Net, inter-sample relational reasoning loss  $L_{inter}$ , the intra-sample relational reasoning loss  $L_{intra}$ , and the external variable relational reasoning loss  $L_{tmp}$  in SpecRR-Net.

Figure 3

See image above for figure legend

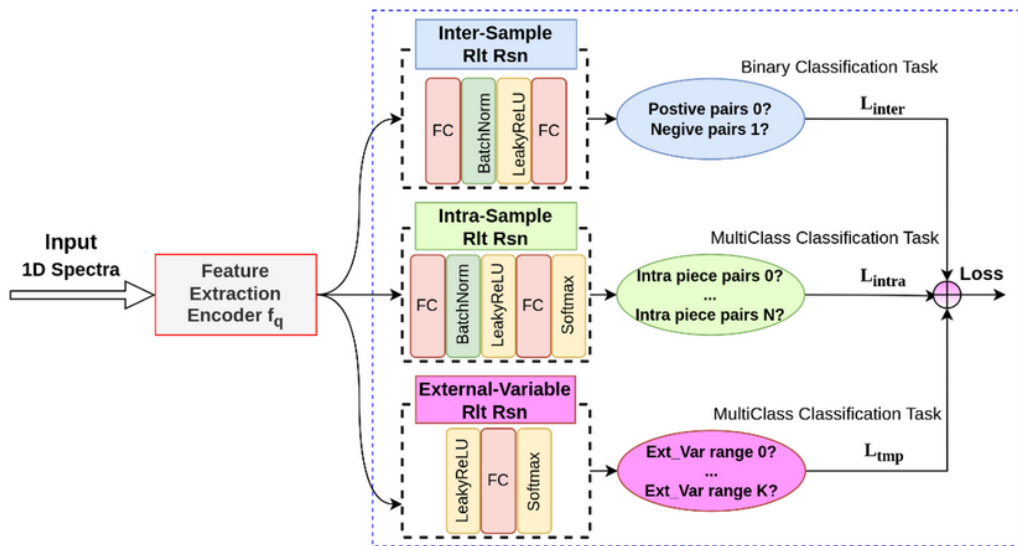


Figure 4. Illustration of the self-supervised relational reasoning sub-network (SpecRR-Net). It mainly consists of inter-sample, intra-sample, and external-variable relational reasoning modules. In this cartoon,  $L_{inter}$  represents inter-sample relational reasoning loss,  $L_{intra}$  represents intra-sample relational reasoning loss,  $L_{tmp}$  represents the external variable relational reasoning loss.

Figure 4

See image above for figure legend

Classification result of SpecRRMoco-Net with 2.7% labels

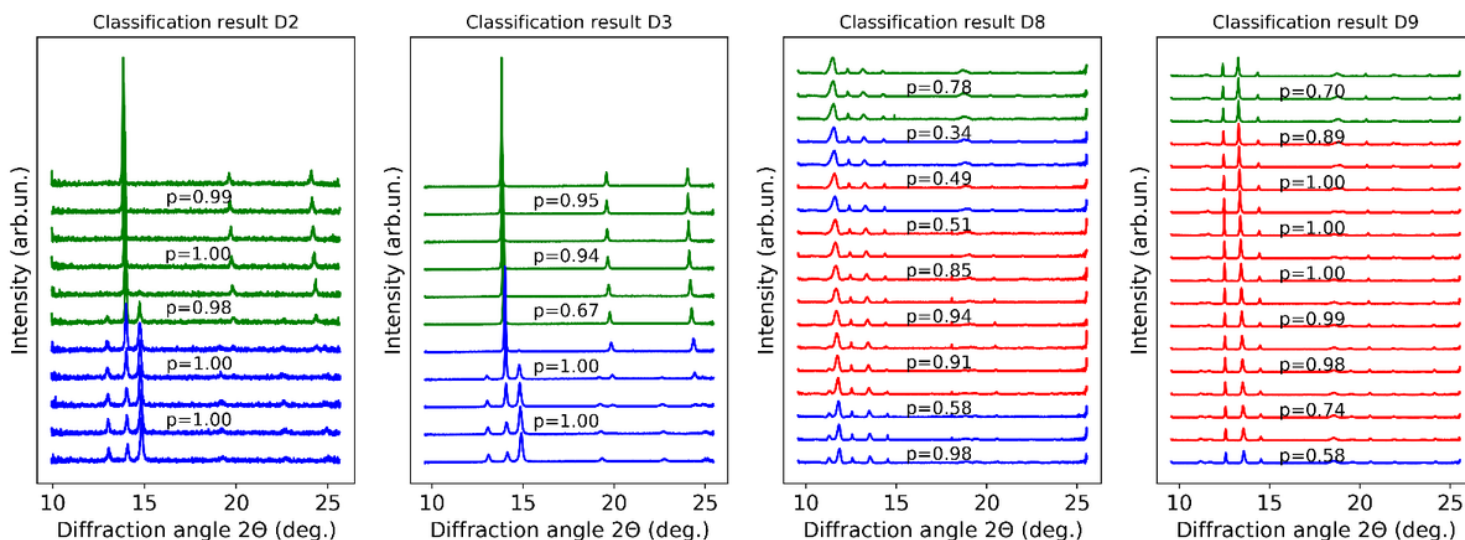


Figure 5

Classification results of the proposed SpecRRMoco-Net on the experimental spectral data with 2.7% labeled data. Each column corresponds to a different data set.

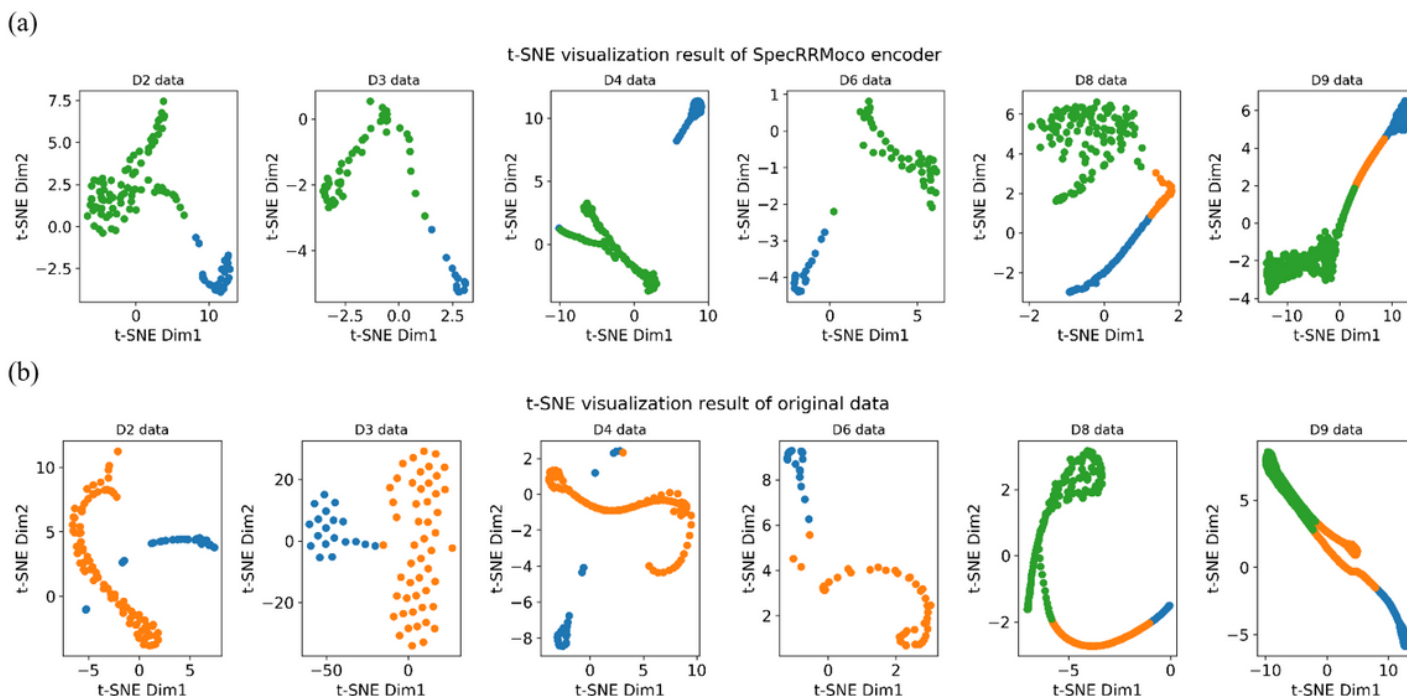
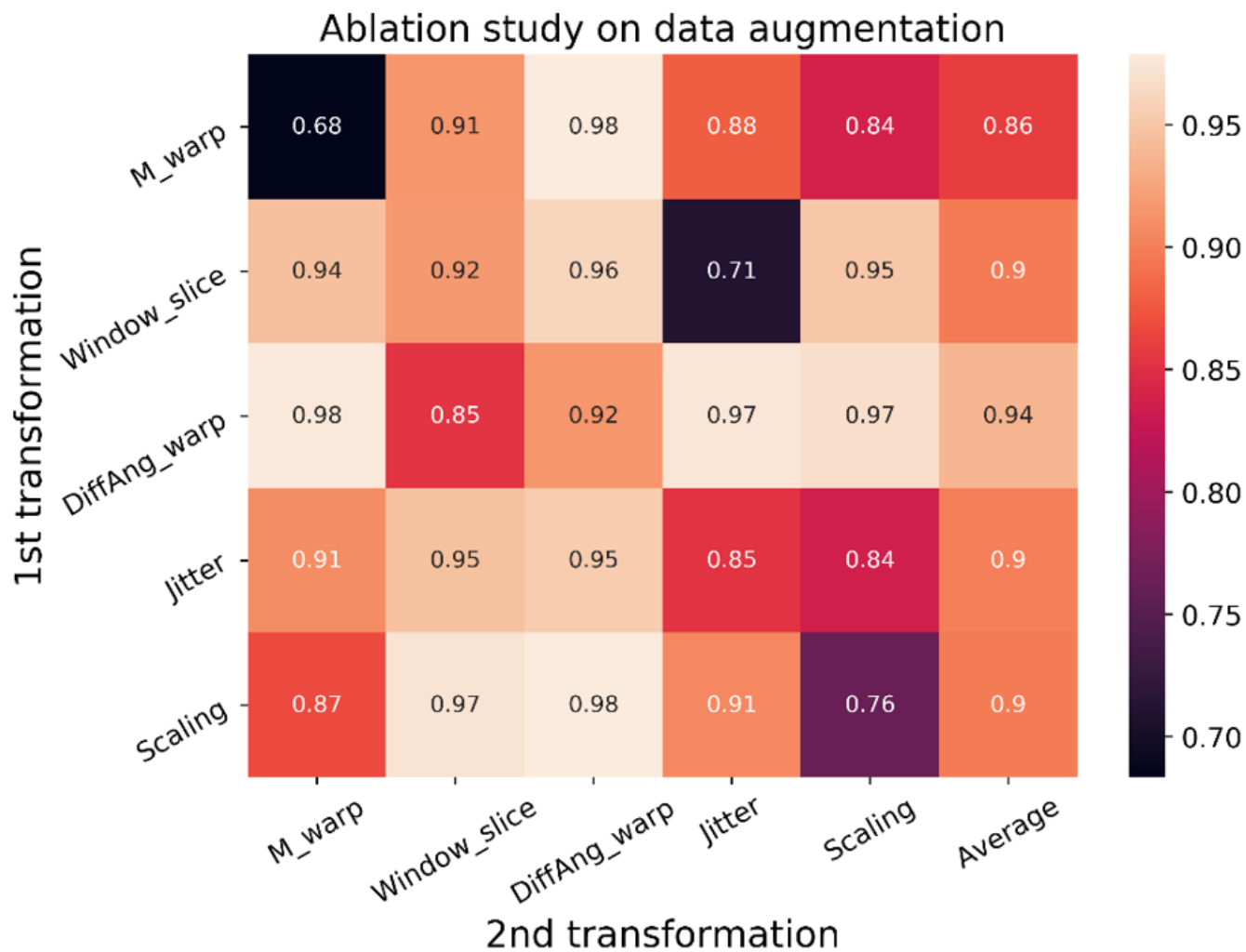


Figure 6

(a) t-SNE visualization of the embedded features after the SpecRRMoco-Net encoder. Class labels were assigned by a linear classifier on top of the embedding representation learned from the SpecRRMoco-Net encoder using 2.7% labeled data. (b) t-SNE visualization of the original data sets clustered by the Spectral Clustering method. In this case, class labels were assigned by applying spectral clustering directly to the original example dataset.



**Figure 7**

Ablation study on data augmentation techniques. Results for magnitude warping (M\_warp), window slicing (Window\_slice), diffraction angle warping (DiffAng\_warp), jitter and scaling are reported. The last column corresponds to the average results over each row (for each data augmentation technique). In addition to this, diagonal elements indicate the use of only one data augmentation technique, while other non-diagonal entries indicate the combination of two data augmentation techniques. The color scale represents the classification accuracy.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SMSelfSupervisedClassificationPaper20230217.docx](#)