# Dynamic Spectrum Matching with One-shot Learning

Jinchao Liu[a], Stuart J. Gibson[a,b], James Mills[c], Margarita Osadchy[d]

[a] *VisionMetric Ltd, Canterbury, Kent, UK. E-mail: liujinchao2000@gmail.com*
[b] *Department of Physical Science, University of Kent, Canterbury, Kent, UK. E-mail: S.J.Gibson@kent.ac.uk*
[c] *Sandexis Ltd, Sandwich, Kent, UK. E-mail: james.mills@sandexis.co.uk*
[d] *Department of Computer Science, University of Haifa, Mount Carmel, Haifa, Israel. E-mail: rita@cs.haifa.ac.il*

**Abstract**

Convolutional neural networks (CNN) have been shown to provide a good solution for classification problems that utilize data obtained from vibrational spectroscopy. Moreover, CNNs are capable of identifying substances from noisy spectra without the need for additional preprocessing. However, their application in practical spectroscopy is restricted due to two reasons. First the effectiveness of classification using CNNs diminishes rapidly when only a small number of spectra per substance are available for training (which is a typical situation in real applications). Secondly, to accommodate new, previously unseen, substance classes the network must be retrained which is computationally intensive. Here we address these issues by reformulating a multi-class classification problem with a large number of classes to a binary classification problem for which the available data is sufficient for representation learning. Hence, we define the learning task as identifying pairs of inputs as belonging to the same class or different classes. We achieve this using a Siamese convolutional neural network. A novel sampling strategy is proposed to address the imbalance problem in training the Siamese network. The trained network can classify samples of previously unseen substance classes using just a single reference sample (termed as one-shot learning in the machine learning community). Our results on three independent Raman datasets demonstrate much better accuracy than other practical systems to date, while allowing effortless updates of the system's database with new substance classes.

*Keywords:* Spectrum Matching, Siamese Network, One-shot Learning, Convolutional Neural Networks

## 1. Introduction

Raman spectroscopy is used for the identification and quantification of solids (particles, pellets, powers, films, fibers), liquids (gels, pastes) and gases. The technique relies upon the inelastic scattering of monochromatic light, caused by interactions with molecular vibrations. A "molecular fingerprint" of a substance can therefore be obtained in the form of a spectrum comprising peaks that are characteristic of its chemical composition. Since it provides fast, non-contact, and non-destructive analysis, Raman spectroscopy has a wide range of applications in a variety of industries and academic fields.

Pattern recognition methods can be used for automatic identification of substances from their Raman spectrum. However, most of the pattern recognition methods require preprocessing of the data. Due to this limitation, a standard pipeline for a machine classification system based on Raman spectroscopy includes preprocessing in the following order: cosmic ray removal, smoothing and baseline correction. Additionally, the dimensionality of the data is often reduced using principal components analysis (PCA) prior to the classification step.

It was shown in [1], that Raman Spectra classification can be achieved successfully using convolutional neural networks (CNNs). There are three main benefits of using CNNs for vibrational spectra classification. Firstly, a CNN can be trained to remove baselines, extract good features, and differentiate between spectra from a large number of classes in an integrated manner within a single network architecture. Thus it removes the need for preprocessing the signal. Secondly, CNN has been shown to achieve significantly better classification results than all previous methods [2, 3, 4, 5, 6, 7, 8, 1]. Thirdly, the classification is very efficient in terms of computation time.

In addition to these factors, a practical spectrum matching system should be enable spectra to be added or removed from a database dynamically in real time, without the additional overheads associated with retraining the underlying classifier. Some substances might have only few or even a single sample per class. Thus a practical system should be able to provide accurate classification even when the training data is sparse. In this context, applying CNN is problematic. First, adding a new class to a CNN requires a change in architecture of the network (or at least its last layers) and retraining of the

network[1] which is computationally intensive and therefore time consuming. Second, training CNN requires many labelled samples, while using one or a few training samples per class dramatically degrades the accuracy of CNN compared to a fully trained CNN or a simpler classifier with a small training set.

Previous methods, that provide the capability of dynamically updating the reference set that contains one or few samples per substance, have tended to use very simple pattern matching algorithms. Typically, commercial systems return a short list of candidate substances ranked according to their similarity with the query spectrum according to the relative magnitudes of their hit quality index (HQI) scores. Different metrics have been used for HQI including Euclidean distance, correlation, and the cosine of the angle between two spectra. The cosine similarity metric has also previously been combined with a nearest neighbour classifier [9, 10]. Variations on these metrics include assigning greater weight to particularly discriminating peaks, or eliminating peaks that only occur in the query spectrum on the assumption that they are due to impurities that are of no interest. The HQI value can be affected by artefacts due to baseline and purge problems and the presence of additional peaks caused by sample contamination and is therefore susceptible to misinterpretation. Databases shipped with commercial vibrational spectroscopy instrumentation sometimes contain records for substances that simply list the positions and intensities of the peaks contained within the spectra and these could be determined from theoretical models (conversely our application is concerned with matching to reference spectra obtained empirically). Although reducing the data to peak positions allows queries to be run quickly, peak width can be important for interpretation [11]. Multi-scale methods make use of the structures of individual peaks [12, 13, 14]. Where a sample contains an unknown mixture of substances, probabilities for the presence of each component may be obtained by, for example, a generalized linear model [15] or reverse searching using non-negative least squares [16]. A reverse search ignores peaks that occur in the query spectrum but not in the reference spectrum contained in the library/database.

To summarize, previous methods for spectra matching do not extract

[1] A more efficient method is to retrain the last layers and only fine tune the rest of the weights. However, the success of this approach depends on the similarity of the new class to the existing ones.

salient features from the data and are therefore highly susceptible to noise and require preprocessing of the raw signal for good matching performance. Hence, CNN is advantages over these methods in providing accurate, efficient, and fully automated classification of spectra. However, 1) it requires large training sets and 2) retraining when a substance is added or deleted from the system. In this paper, we present a system that enjoys all the benefits of CNN and solves these two problems. Our experiments show that the proposed method can perform an accurate classification of spectra even in cases where the number of classes is large and with a single or a handful of samples per class. Moreover, it allows new classes to be added or existing classes to be removed from the model in real time with no additional effort.

## 2. Materials and Methods

### *2.1. Our Approach*

Our approach tackles the two issues of CNNs that restrict their use in practical systems: degradation in accuracy due to lack of training data and inability to perform online updates of the database.

**Training with limited data:** Here the issue of limited available data for training the CNN is addressed by reformulating an n-way classification task into a binary task in which pairs of inputs are assigned to the same class or different classes. For this binary problem, applied to a large number of classes, even a small number of samples per class would result in a large number of training pairs (we provide further details in Section 2.3). To learn the resulting binary problem we use a special architecture, referred to as Siamese network [17], which has been used to determine if pairs of inputs originate from the same class in a number of domains (e.g., RGB images [18, 19, 20], NIR images [21], speech [22] and text [23]).

**Online updates:** The Siamese network architecture can be viewed as a combination of a non-linear mapping for extracting features from the input pairs with a weighted metric for comparing the resulting feature vectors. The mapping and the metric are learned in an integrated manner using gradient-based learning. When a Siamese network is trained on many classes, the resulting features and the corresponding learned metric are capable of generalizing beyond the classes seen in training. Thus it can be used for learning to process data from unseen classes using a single training sample, termed one-shot learning. Previous work showed the merit of using Siamese networks for one-shot learning in character recognition and object classification

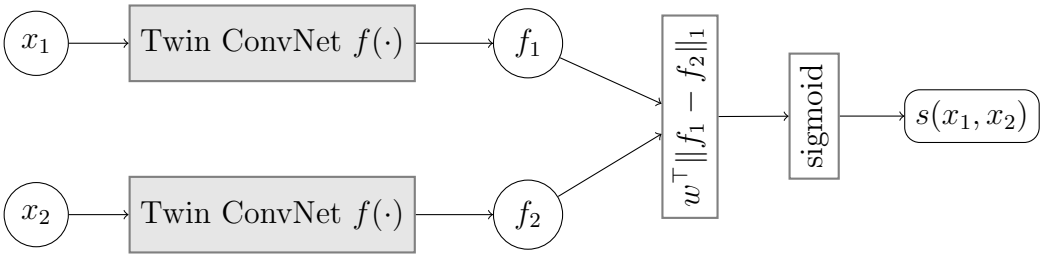



Figure 1: Diagram of a Siamese network with a convolutional neural network as its twin network. $x_1$ and $x_2$ are two samples to compare, $f_1$ and $f_2$ denote their features extracted by the twin CNN. The metric in the feature space has chosen to be weighted $L_1$ which is learnable by adjusting $w$[24]. Finally the network outputs a similarity measure $s(x_1, x_2) \in [0, 1]$.

(e.g. [18, 24]). We propose applying one-shot learning, using a Siamese network, for spectra classification and use it to build a dynamic classification system that enables online updating of a spectrum database (without retraining the system). Specifically, for an n-way classification problem, we use a single reference sample per class, including the new classes that were not previously represented during the training of the Siamese network, and map the reference samples and the test sample to the feature space via the CNN part of the the Siamese network. Then, the nearest neighbour rule is applied using the learned similarity metric for classifying the test sample. If more samples are available per class, the comparison can be extended to k-nearest neighbors. One can also perform ranking or any other analysis of distances between the test sample to the reference set.

### *2.2. Siamese Network for One-Shot Learning*

Siamese networks [18] consist of two (twin) networks that have exactly the same structure and identical weights. The architecture of the Siamese net used in this work is shown in Figure 1. We implemented the twin networks using the CNN architecture shown in Figure 2. The twin network maps an input spectrum to a feature space using the same mechanism as is employed for classification using a single CNN [1] and thus enjoys all the benefits of CNN as detailed above. The CNN architecture includes six blocks, each with a convolutional layer, followed by batch normalization, LeakyReLU nonlinearity, and max-pooling. The number of feature maps is decreased in every second layer. The outputs of the last block are concatenated and flattened to form a feature vector, which is used as an input to the metric learning

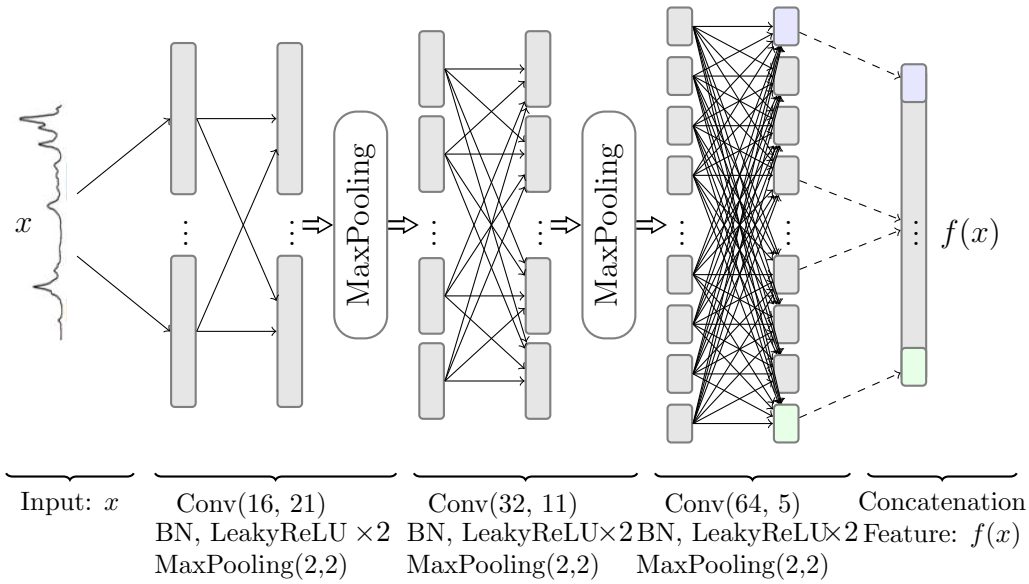



Figure 2: Diagram of convolutional neural networks that are variants of LeNets[25]. This was used in our work as the twin network in Siamese networks. BN stands for batch normalization. Conv(m, n) stands for a convolutional layer with m neurons/filters of kernel size n. Maxpooling($\sigma, s$) denotes a MaxPooling layer with kernel size $\sigma$ and stride $s$.

part of the Siamese net. Further details of the CNN architecture are shown in Figure 2. We note that all the building blocks including convolutional layers and max-pooling are one dimensional as we represent spectra as 1D sequences of intensity at each wavenumber.

The Siamese network computes the following function:

$$s(\mathbf{x}_i, \mathbf{x}_j) = \sigma(-\mathbf{w}^\top \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|_1) \tag{1}$$

where $\mathbf{x}_i, \mathbf{x}_j$ are a pair of positive or negative samples, $\sigma$ is the sigmoid activation function. $f$ is a non-linear transform realized by a convolutional network (twin network) and $\mathbf{w}$ are the trainable weights that can be viewed as a metric in the feature space. We use a binary cross-entropy loss function to train the Siamese network

$$\mathcal{L} = -\sum_{n=1}^{N} \Big[ y_n \log s(\mathbf{x}_{n_i}, \mathbf{x}_{n_j}) + (1 - y_n) \log(1 - s(\mathbf{x}_{n_i}, \mathbf{x}_{n_j}) \Big] \tag{2}$$

where N is the total amount of positive and negative pairs.

### *2.3. Imbalanced Positive and Negative Pairs: A Sampling Strategy*

One of the core problems associated with training a Siamese network is generating negative and positive pairs efficiently and sufficiently for the purpose of distinguishing both similar and dissimilar samples. Suppose there are $N$ classes, each of which has $M$ samples, the total number of positive and negative pairs are $M(M-1)N/2$ and $M^2N(N-1)/2$, respectively. In applications like mineral recognition, in which there are typically hundreds of different minerals, but only a handful of spectra available for each, there will be considerably more negative than positive pairs i.e. for $M \sim 10$ and $N \sim 100$, there will be roughly $5K$ positive and $500K$ negative pairs. If we feed this ratio of training pairs into the network without proper countermeasures, positive pairs may be dominated by negative pairs during training and will therefore be under-represented in the model. As a result, the network is likely to be biased towards distinguishing dissimilar pairs. An extreme case would be a classifier that only reports negative responses and fails to learn variance within the same class or similar samples.

A common way of dealing with an imbalanced dataset is to under-sample the majority class i.e. negative pairs in this case, or over-sample the minority class i.e. positive pairs, or to combine both these strategies [26]. Here, we propose using a Bootstrapping-based strategy as follows: assume that we have $S_m$ positive pairs and $S_n$ negative pairs where $S_m \ll S_n$, for each iteration. We then sample $S_m$ positive pairs and an equal number of negative pairs, both with replacement. We repeat a number of iterations until the Siamese network has been trained sufficiently. It should be noted that it is crucial to sample the positive pairs with replacement, instead of generating all the positive pairs deterministically, to prevent overfitting to positive pairs.

### *2.4. Training Convolutional Siamese Nets*

It is common practice to increase the number of training samples by augmenting the real samples with the synthetic ones. Here, we employed data augmentation technique, proposed in [1], which consists of three steps. First, each spectrum is shifted left or right a few wave numbers randomly. Then, random noise, proportional to the magnitude at each wave number, is added. Lastly, the augmented data is created as linear combinations of all spectra, belonging to the same substance, with random coefficients.

The Siamese network was trained using the Adam algorithm[27] (a variant of stochastic gradient descent) for 50 iterations. The learning rate was

reduced by half every 10 iterations. Xavier initialization was used to initialize the convolution layers. We applied early stopping to prevent overfitting. Training was performed on a single NVIDIA GTX-1080 GPU.

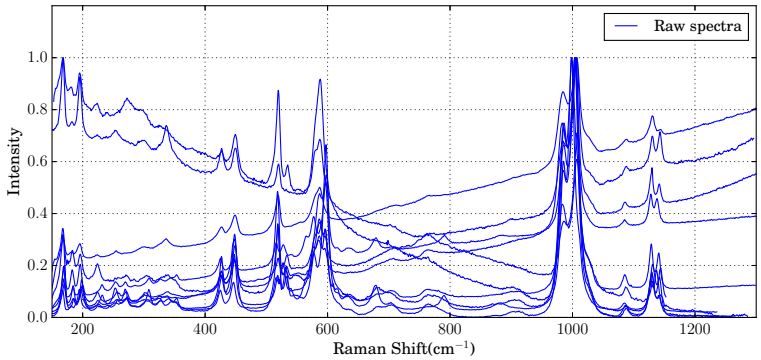
### *2.5. Raman Spectroscopy Datasets*

We tested the proposed method on three Raman spectroscopy datasets, two public available mineral datasets (RRUFF [28] and UNIRP [29]) and one varied chemical dataset (CHEMK). RRUFF is the largest publicly available mineral database. In our experiment, we used raw (uncorrected) spectra of 512 minerals and around 1700 spectra in total. UNIRP is a smaller mineral database, containing 107 minerals and 163 spectra in total. The chemical dataset, referred to as CHEMK in our paper, contains spectra of 123 chemicals and around 500 spectra in total.

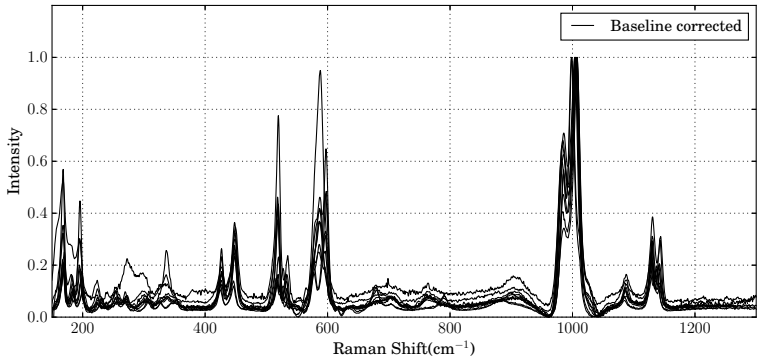
We note that the RRUFF dataset contains only spectra that were not baseline corrected, while UNIPR and CHEMK datasets contain both baseline corrected and raw spectra. In order to investigate the performance of the tested methods on baseline corrected spectra, we applied a widely-used baseline correction technique, *asymmetric least squares* [30], to produce a baseline-corrected version of corresponding datasets.

## 3. Results and Discussion

Our experimental worked was organised into two parts. In the first part ( Section 3.1) we evaluated our method for the real-world application setting in which online changes are made to the database (the reference set). The tests show a significant advantage of the proposed method over the established matching methods on three different datasets. In the second part of the experimental work (Section 3.2) we compared the performance of the Siamese network with a standard CNN approach [1] for spectrum classification (which assumes that all the classes are known during training and thus does not allow for online updates). The results of this experiment show that the proposed method is comparable to CNN in its ability to extract good features (which all previous matching methods lack). Finally, we use a visualization technique (shown in Section 3.3) to illustrate the ability of the proposed method to map samples of different classes to non-overlapping clusters. This explains its excellent performance in spectra matching.




(a) Raw spectra where baselines can be observed.




(b) Baseline corrected by asymmetric least squares

Figure 3: Spectra of a mineral, *hydroxylherderite*, from RRUFF raw database and corresponding baseline corrected spectra by asymmetric least squares.

### *3.1. Spectrum Matching*

In spectrum matching the updates of the reference database with the new classes must be done online, without any additional training. Thus the system should be capable of processing instances of new classes that were not included in training. The following experiment compares the performance of the proposed method with three other popular methods for matching against previously unseen classes.

#### *3.1.1. Evaluation Protocol*

We investigated the performance of the proposed method on unseen classes using the following protocol: We split all classes into non-overlapping sets for training, validation, and test. We used 50% of the classes for training, 10% for validation, and 40% for testing. We trained the Siamese network using all samples in the training set as discussed in Section 2.3. We validated the results for early stopping on pairs produced from the samples in the validation set. During test time, we picked at random a single sample from each class in the test set to form a reference set. We performed identification of all spectra in the test set by matching them to the reference set. Using a randomized partitioning of the data, we repeated both training and test procedures several times to obtain statistically reliable results.

UNIPR dataset is much smaller than RRUFF and CHEMK. Specifically, it has many classes with a single sample. A matching test requires at least two samples: one reference and one test, thus single sample classes cannot be used in the test time. To this end, we used a different evaluation protocol based on a fixed split of classes for UNIPR tests. We sorted all the minerals according to the number of spectra they contain in descending order and selected the first 20 mineral species for testing, the second 20 for validation and the rest for training (as training can be done with a single sample per class). Following this protocol, we can make sure that all the test classes have at least two spectra per class. Note that, since the training classes have fewer samples (usually one spectrum per mineral) this protocol yields a harder problem to solve than the one based on a random split.

We compared our method with three previous matching techniques: nearest neighbour (NN) with $L_2$ distance, nearest neighbour with cosine similarity and large margin nearest neighbour (LMNN). Nearest neighbour with the Euclidean distance and cosine similarity have been widely used in commercial software for spectrum matching and were therefore included for comparison.

| Dataset | Signal Type | NN($L_2$) | NN(cosine) | LMNN[31] | Siamese Net |
|---|---|---|---|---|---|
| RRUFF | Raw | 0.461±0.046 | 0.525±0.036 | 0.725±0.041 | 0.901±0.014 |
| | Preprocessed | 0.802±0.033 | 0.833±0.030 | 0.818±0.029 | 0.886±0.032 |
| CHEMK | Raw | 0.827±0.047 | 0.879±0.037 | 0.882±0.043 | 0.972±0.024 |
| | Preprocessed | 0.796±0.030 | 0.838±0.026 | 0.878±0.047 | 0.974±0.024 |
| UNIPR | Raw | 0.630 | 0.723 | 0.803 | 0.965 |
| | Preprocessed | 0.724 | 0.835 | 0.838 | 0.956 |

Table 1: One-shot classification accuracy of convolutional Siamese nets and other compared methods with and without baseline correction.

LMNN is a popular metric learning method that determines a Mahanalobis distance for $k$-nearest neighbour classification ($k$NN). A linear transform of the input space is learned such that the $k$-nearest neighbors of a sample in the training set share the same class label with the sample, while samples from different classes are separated by a large margin. In short, LMNN learns a linear transform, which is particularly beneficial for $k$NN classification. Hence the compared methods include from widely used non-learning, limited-learning to powerful meta-learning methods(Siamese network) [2] .

*3.1.2. Results and Analysis*

The results, summarized in Table 1, indicate that our proposed method outperforms all other tested methods by a large margin and removes the need for data preprocessing. In the following, we analyze these results in detail.

**RRUFF Dataset**: On the *raw data* NN with either $L_2$ or cosine similarity performed poorly, with a low accuracy rate of $\sim 0.5$. LMNN achieved better results, with an accuracy of 0.725. The Siamese network significantly outperformed all tested methods and produced the highest rate of 0.901. On the *baseline corrected* (i.e. preprocessed) data, all previous methods performed much better. The Siamese network achieved 0.886, which is lower than on the raw data but still significantly better than the other tested

[2]Meta-learners is a term used in meta learning and one shot learning for describing machine learning methods which can learn e.g. cross-domain knowledge and generalize to unseen tasks. Siamese networks can be regarded as a simple yet powerful meta- learning method.

methods. These results support the preconception that convolutional networks achieve the best integrated preprocessing and feature extraction of raw spectra. A similar conclusion was drawn in [1] where the end-to-end training of a CNN for Raman spectra classification produced superior results on raw data compared to pipeline approaches.

It is worth noting that on the raw data, LMNN achieved better results than NN with both $L_2$ and cosine similarity (as shown in Table 1). LMNN is capable of learning a linear transform to facilitate the subsequent nearest neighbour classification. Such a simple transform is much more limited than the one learned by the Siamese Network, but it does result in a modest improvement when applied to raw, unprocessed, spectra. An example in which a linear transform can provide effective preprocessing is depicted in Figure 3(a). We can see that rotating the spectra clockwise or counterclockwise respectively, which is a linear transform, would correct the baselines to a some degree.

The result of LMNN on the preprocessed data was comparable to NN with the $L_2$ distance metric and worse than NN with cosine similarity. This further supports the limited ability of LMNN to learn preprocessing.

**CHEMK and UNIPR Datasets**: Siamese net significantly outperformed the compared methods on CHEMK and UNIPR datasets. Specifically, the accuracy was increased by more than 10% compared to the second best method LMNN and the widely used cosine similarity method. For the CHEMK dataset, baseline correction worsened the accuracy of all methods except for the Siamese net. This may imply that the standard baseline correction techniques may also inadvertently remove some information content from spectra that is useful for classification. We emphasize that learned preprocessing, as in Siamese Net, is tuned to the data and thus does not fail.

### *3.2. Comparison to Multi-class CNN*

The following set of experiments aims to test the ability of the proposed method to extract good features. To this end we compare its classification performance to the state-of-the art Raman spectra classification method based on CNN [1] in a multi-class classification, where unseen samples are classified into learned classes. In addition, our experiments test how much these methods rely on data augmentation (customary in CNN training).

| Dataset | Methods | Data Augmentation | Raw | Preprocessed |
|---|---|---|---|---|
| RRUFF | ConvNets[1] | No♣ | 0.705±0.031 | 0.779±0.015 |
| | | Yes | 0.933±0.007 | 0.920±0.008 |
| | Siamese Nets | No | 0.850±0.029 | 0.842±0.035 |
| | | Yes | 0.921±0.010 | 0.919±0.007 |
| CHEMK | ConvNets[1] | No | 0.764±0.027 | 0.758±0.032 |
| | | Yes | 0.891±0.012 | 0.877±0.016 |
| | Siamese Nets | No | 0.839±0.026 | 0.805±0.029 |
| | | Yes | 0.903±0.025 | 0.883±0.021 |

Table 2: Classification accuracy on trained classes of convolutional Siamese nets and CNNs on the RRUFF mineral and CHEMK chemical datasets, with or without baseline correction.

*3.2.1. Evaluation Protocol*

We applied the Siamese network as described in Section 3.1, but with the reference set composed of the known classes (used to train the Siamese network). For this set of experiments, we followed the test protocol[1], which randomly selects one sample from each class to form a test set. The remaining samples were used for training and validation. The process was repeated a number of times. Due to lack of training data (many classes with a single spectrum per mineral) we did not include UNIPR dataset in this set of experiments.

*3.2.2. Results and Analysis*

The results are summarized in Table 2. We can see that when data augmentation is used in training, the classification accuracy of the Siamese network is comparable to CNN on the preprocessed spectra and only slightly lower than CNN on the raw data. This indicates that the proposed method is able to extract very good features, comparable to the state-of-the-art CNN.

Additionally, one can consider the Siamese Network as an alternative to CNN for a large-scale classification of spectra when the training data is limited. To test this hypothesis, we investigated the effect of data augmentation on convolutioinal Siamese networks and on CNNs.

Table 2 shows that the convolutional Siamese network trained with *no data augmentation* is able to achieve a good accuracy on both raw and preprocessed spectra. A trained CNN network with no augmentation, was heavily

overfitted and showed much lower success rates than the Siamese network. To reduce overfitting of CNN we tried halving its size by keeping only one copy of each block (instead of two). We tested the reduced CNN on RRUFF dataset and found that reducing the size of CNN resulted in a significant loss of accuracy compared to the original network trained on augmented data, specifically, 70.5% on raw spectra and 77.9% on preprocessed spectra. Furthermore, the reduced CNN attained higher accuracy on the preprocessed data than on the raw spectra, which is contrary to the results of the original CNN trained on the augmented data. This suggests that the reduced CNN is not large/deep enough to learn a baseline correction well.

These results confirmed that the convolutional Siamese network is a better alternative than multi-class CNN in classification tasks with a large number of classes but a small number of samples per class, and when synthetic augmentation of the data is not possible. The success of Siamese net in this experiment can be explained by the fact that the number of pairs obtained from a data set comprising many classes with a small number of samples in each, is large enough to prevent significant overfitting.

### *3.3. Visualization*

Our method learns to map the input samples to non-overlapping clusters which facilitates the matching. To verify this claim, we used a visualization technique that maps high-dimensional features learned by the twin networks in the Siamese net to low-dimensional space while preserving the pairwise distances.

We used t-distributed stochastic neighbour embedding (t-SNE)[32] to visualize the feature space learned by the Siamese network. To avoid clutter in the visualization, we reduced the number of unseen classes by removing the minerals with fewer than three samples. The results depicted in Figure 4 show that the projected samples cluster by mineral type.

## 4. Conclusion and Future Work

In this paper, we have proposed a one-shot learning solution based on convolutional Siamese networks to realize a dynamic spectrum matching system which is capable of classifying both seen and unseen classes accurately. Importantly, for unseen classes/substances, the proposed system requires as few as one example per class to achieve accurate classification which enables classes/substances to be added into the model dynamically. We validated

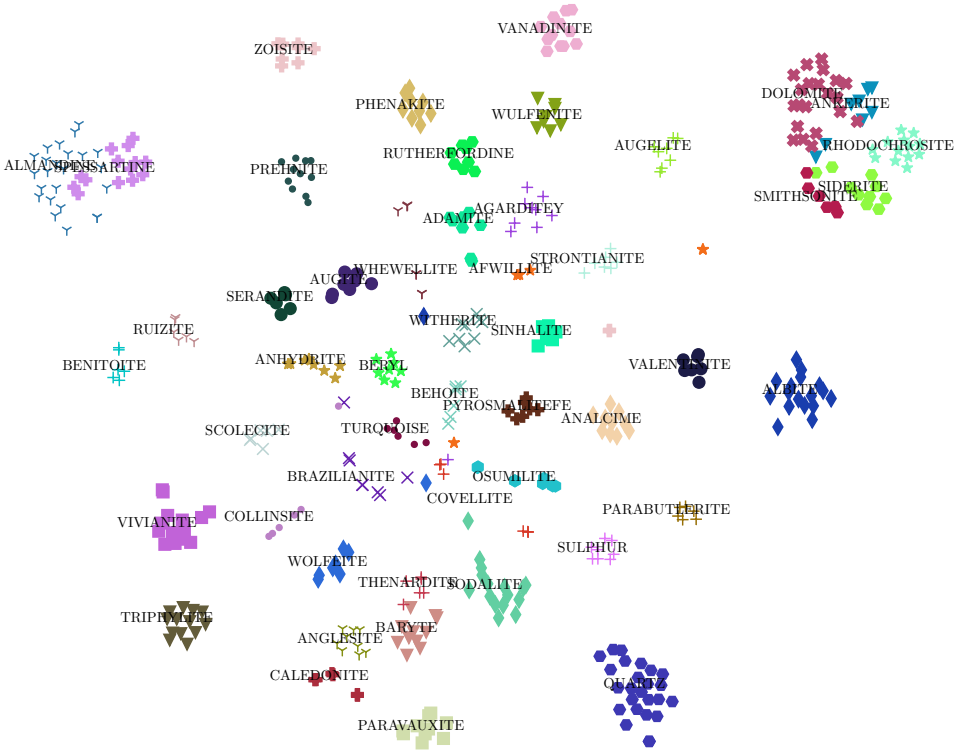



Figure 4: t-SNE projection of the transformed features of some test samples. Classes are highlighted in different colors.

the feasibility and effectiveness of our method on the largest public available mineral dataset. Although we demonstrated our method for Raman spectroscopy, it can also be applied to other kinds of spectroscopy and other applications such as computational chemistry, drug discovery, and health care. The approach is particularly applicable for cases where the data available for training is limited and/or frequent updates of the reference database are required. The one-shot solutions are not limited to Siamese Networks. Some different one-shot architectures have previously been applied to related fields such as drug discovery (see also Altae-Tran et al. [33]). We anticipate further development of our work to include an adaptation and evaluation of different one-shot learning methods in spectroscopy.

## 5. Acknowledgements

We thank IS-Instruments for providing the CHEMK dataset used in this work.

## 6. References

[1] J. Liu, M. Osadchy, L. Ashton, M. Foster, C. J. Solomon, and S. J. Gibson, "Deep convolutional neural networks for raman spectrum recognition: a unified solution," *Analyst*, vol. 142, pp. 4067–4074, 2017.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, pp. 1097–1105, 2012.

[3] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 3304–3308, Nov 2012.

[4] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds.), pp. 3104–3112, Curran Associates, Inc., 2014.

[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Computer Vision and Pattern Recognition*, 2014.

[6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2017.

[7] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to End Learning for Self-Driving Cars," *ArXiv e-prints*, Apr. 2016.

[8] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen, J. Chen, J. Chen, Z. Chen, M. Chrzanowski, A. Coates, G. Diamos, K. Ding, N. Du, E. Elsen, J. Engel, W. Fang, L. Fan, C. Fougner, L. Gao, C. Gong, A. Hannun, T. Han, L. Johannes, B. Jiang, C. Ju, B. Jun, P. LeGresley, L. Lin, J. Liu, Y. Liu, W. Li, X. Li, D. Ma, S. Narang, A. Ng, S. Ozair, Y. Peng, R. Prenger, S. Qian, Z. Quan, J. Raiman, V. Rao, S. Satheesh, D. Seetapun, S. Sengupta, K. Srinet, A. Sriram, H. Tang, L. Tang, C. Wang, J. Wang, K. Wang, Y. Wang, Z. Wang, Z. Wang, S. Wu, L. Wei, B. Xiao, W. Xie, Y. Xie, D. Yogatama, B. Yuan, J. Zhan, and Z. Zhu, "Deep speech 2 : End-to-end speech recognition in english and mandarin," in *Proceedings of The 33rd International Conference on Machine Learning* (M. F. Balcan and K. Q. Weinberger, eds.), vol. 48 of *Proceedings of Machine Learning Research*, (New York, New York, USA), pp. 173–182, PMLR, 20–22 Jun 2016.

[9] S. T. Ishikawa and V. C. Gulick, "An automated mineral classifier using raman spectra," *Comput. Geosci.*, vol. 54, pp. 259–268, Apr. 2013.

[10] C. Carey, T. Boucher, S. Mahadevan, P. Bartholomew, and M. Dyar, "Machine learning tools for mineral recognition and classification from raman spectroscopy," *Journal of Raman Spectroscopy*, vol. 46, no. 10, pp. 894–903, 2015.

[11] A. Goss and M. Adams, "Spectral retrieval by fuzzy matching," in *Analytical Proceedings including Analytical Communications*, vol. 31, pp. 23–25, Royal Society of Chemistry, 1994.

[12] Z.-M. Zhang, X. Tong, Y. Peng, P. Ma, M.-J. Zhang, H.-M. Lu, X.-Q. Chen, and Y.-Z. Liang, "Multiscale peak detection in wavelet space," *Analyst*, vol. 140, no. 23, pp. 7955–7964, 2015.

[13] H.-Y. Fu, J.-W. Guo, Y.-J. Yu, H.-D. Li, H.-P. Cui, P.-P. Liu, B. Wang, S. Wang, and P. Lu, "A simple multi-scale gaussian smoothing-based strategy for automatic chromatographic peak extraction," *Journal of Chromatography A*, vol. 1452, pp. 1–9, 2016.

[14] X. Tong, Z. Zhang, F. Zeng, C. Fu, P. Ma, Y. Peng, H. Lu, and Y. Liang, "Recursive wavelet peak detection of analytical signals," *Chromatographia*, vol. 79, no. 19-20, pp. 1247–1255, 2016.

[15] T. Vignesh, S. Shanmukh, M. Yarra, E. Botonjic-Sehic, J. Grassi, H. Boudries, and S. Dasaratha, "Estimating probabilistic confidence for mixture components identified using a spectral search algorithm," *Applied spectroscopy*, vol. 66, no. 3, pp. 334–340, 2012.

[16] Z.-M. Zhang, X.-Q. Chen, H.-M. Lu, Y.-Z. Liang, W. Fan, D. Xu, J. Zhou, F. Ye, and Z.-Y. Yang, "Mixture analysis using reverse searching and non-negative least squares," *Chemometrics and Intelligent Laboratory Systems*, vol. 137, pp. 10–20, 2014.

[17] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a "siamese" time delay neural network," in *Proceedings of the 6th International Conference on Neural Information Processing Systems*, NIPS'93, pp. 737–744, 1993.

[18] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, CVPR '05, (Washington, DC, USA), pp. 539–546, IEEE Computer Society, 2005.

[19] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pp. 1701–1708, 2014.

[20] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, 2015.

[21] C. A. Aguilera, F. J. Aguilera, A. D. Sappa, C. Aguilera, and R. Toledo, "Learning cross-spectral similarity measures with deep convolutional neural networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016.

[22] K. Chen and A. Salman, "Extracting speaker-specific information with a regularized siamese deep network," in *Advances in Neural Information Processing Systems 24* (J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, eds.), pp. 298–306, 2011.

[23] Z. Lu and H. Li, "A deep architecture for matching short texts," in *Advances in Neural Information Processing Systems 26* (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, eds.), pp. 1367–1375, 2013.

[24] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," 2015.

[25] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, pp. 2278–2324, 1998.

[26] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[27] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[28] B. Lafuente, R. T. Downs, H. Yang, and N. Stone, "The power of databases: the rruff project," *Highlights in Mineralogical Crystallography*, pp. 1–30, 2015.

[29] DiFeST, "http://www.fis.unipr.it/phevix/ramandb.php."

[30] P. H. C. Eilers and H. F. Boelens, "Baseline correction with asymmetric least squares smoothing," tech. rep., Leiden University Medical Centre, Oct 2005.

[31] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, June 2009.

[32] L. van der Maaten and G. Hinton, "Visualizing high-dimensional data using t-sne," 2008.

[33] H. Altae-Tran, B. Ramsundar, A. S. Pappu, and V. S. Pande, "Low data drug discovery with one-shot learning," in *ACS central science*, 2017.