Siamese Networks for 1D Signal Identification

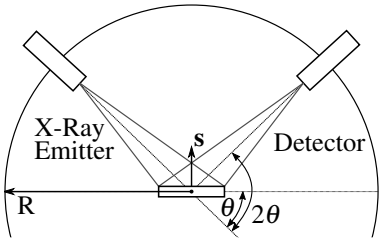
Jan Schuetzke[1], Alexander Benedix[2], Ralf Mikut[1], Markus Reischl[1]

[1] Institute for Automation and Applied Informatics, KIT
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen
E-Mail: jan.schuetzke@kit.edu

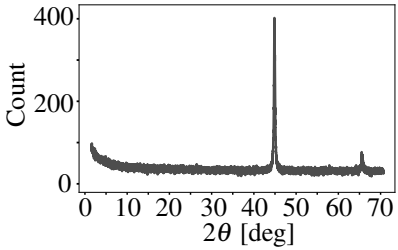
[2] Bruker AXS GmbH
Oestliche Rheinbrückenstraße 49
76187 Karlsruhe
E-Mail: alexander.benedix@bruker.com

Abstract

In material sciences, X-ray diffraction (XRD) or nuclear magnetic response (NMR) are methods to generate one-dimensional signals, describing intensities over an angle or a chemical shift. Each material has a characteristic profile and unknown samples are typically matched to known references. Automatic classification of one-dimensional signal patterns is a non-trivial task due to background noise and varying positions of measured intensities in identical probes. Convolutional Neural Networks prove to be particularly suitable, a limitation, though, is that adding new classes requires retraining. However, continuous discovery of new materials requires possibilities for easy class-extension. Siamese Neural Networks are able to extend data set classes easily and are popular in the field of face recognition, where new faces are constantly added to the database of references. In this paper, we apply Siamese networks to one-dimensional XRD-data for the first time and discuss the opportunities and challenges as well as areas of application. We show that Siamese networks are well suited for the transfer between XRD datasets, achieving an accuracy of 99% for materials not present in the training dataset.



(a) Experimental setup for measuring a XRD scan using the Bragg-Brentano focusing geometry



(b) Resulting XRD scan of an exemplary crystalline sample

Figure 1: Layout and functionality of an instrument for measuring XRD scans and a resulting diffraction pattern

1 Introduction

A typical task in the field of material analysis is to analyze an unknown sample in order to determine the underlying properties or to assign the sample to a known material. For this purpose, a variety of destructive and non-destructive methods exist, which scan the samples with a measuring instrument and then output a one- to multi-dimensional signal, which is subsequently analyzed using software or expert knowledge. Due to the large number of samples to be analyzed and the involvement of experts for the interpretation of the measurement signals, it is especially suitable for this field to utilize automated approaches to speed up the analysis of unknown material samples.

One of the representatives of the methods for the analysis of crystalline samples is the X-ray diffraction (XRD), which utilizes Bragg's Law to deduce the underlying crystal structure of the sample and thus, to determine the corresponding mineral. In the most prevalent technique for measuring XRD signals, called the Bragg-Brentano focusing geometry, the crystalline sample is crushed into a powder and measured using a moving pair of emitter and detector. Figure 1a shows the layout for an instrument that employs the Bragg-Brentano geometry. The powdered sample is placed in the center and X-ray emitter as well as detector are moving upwards on the circular orbit with radius R. The detector measures the intensity of diffracted X-rays subject to incident

angle θ. Figure 1b visualizes the resulting, one-dimensional XRD scan with the measured intensity (count) as a function of the emergent angle, typically described by 2θ [1].

For the XRD scan analysis, each crystalline material forms a distinct diffraction pattern which is subject to the underlying crystal lattice properties. It is of interest where peaks are located and how the intensities are distributed. The X-rays are reflected by atomic planes, so the number and location of peaks depends on the form and shape of the lattice, while the intensity is relative to the comprised atoms. Accordingly, the XRD scans are analyzed by comparing the diffraction pattern to a database of reference minerals. Some natural influences mean that the patterns are not exactly the same, causing small deviations in peak position, shape and intensities [1]. Thus, matching measured and reference patterns is not a trivial task, so professional expertise is required for the manual evaluation of XRD scans.

Similarly, other methods for analyzing unknown material samples, like the proton nuclear magnetic resonance spectroscopy (^{1}H NMR) or infrared spectroscopy, also feature a one-dimensional signal that is characterized by peak locations, shape and intensities. Consequently, a variety of algorithms for the automated analysis have been developed, with a focus on machine and deep learning models in recent years, which gained great popularity especially in the field of image analysis [2, 7, 4]. In many applications, neural networks are employed as classifiers, which have a clear disadvantage for the use with material science data: Once trained for a certain dataset, no further materials can be added without the network (or at least some of its layers) having to be retrained.

A special, extendable type of neural networks are the so-called Siamese networks, which do not learn any classification rules, but distinguish representatives of one class from those of other classes by means of a distance function. Thus, the network deduces the features of two inputs to be compared in two identical, parallel processing lines, allowing the references to be constantly changed and extended, if the type of the calculated features remains the same. Typically, this type of neural network is used in the field of face recognition, where the inputs in the form of faces are always somewhat-similar but the

network extracts characteristic features to distinguish dissimilar from similar faces. Likewise, one-dimensional data such as XRD or NMR scans, which consist of characteristic peaks at specific locations, are potentially suitable for use with Siamese networks.

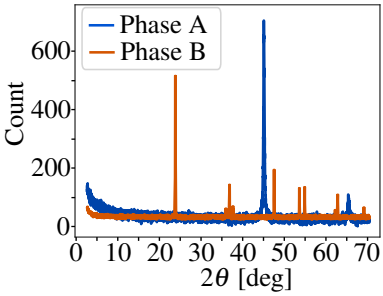
Accordingly, we apply Siamese neural networks for the use with one-dimensional data for the first time, using XRD data as an example. In contrast to regular neural network classifiers, which permanently learn the classification rules in the training phase and are therefore not extendable, the scan to be evaluated can be compared to a dynamically composed set of references for Siamese Networks. Hence, we

1. train a classifier and Siamese network on one dataset
   to compare the baseline performance,
2. apply the trained Siamese network to a second dataset without retraining,
3. retrain the classifier for the second dataset using transfer learning.

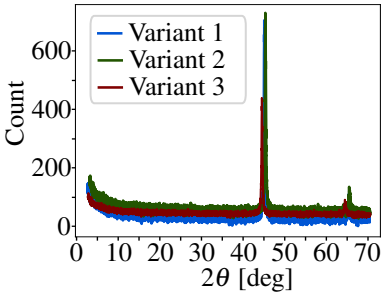
2 Related Work

The XRD scans of distinct phases differ in the number of peaks and the corresponding positions, as demonstrated in Figure 2a. Between diffraction patterns of the identical material, natural influences cause small deviations of peak position, shape and intensities, which is visualized in Figure 2b. For example, stress or small deviations of the lattice lead to a shift of the peaks, varying crystallite sizes cause to wider, flatter peaks and non-ideal preparation of the samples lead to changes of the intensity ratios [1]. The classifiers receive the XRD scans as input and must learn to distinguish between variations within a phase and diffraction patterns of different materials.

Naturally occurring crystalline materials, such as ores, are usually not pure minerals, but rather contain one major phase and small amounts of minor phases. Thus, numerous works employed the Non-Negative Matrix Factorization (NMF) algorithm, which tries to describe the measured signal by combining several components with different proportions [2]. Through a supervised training



(a) XRD scans of exemplary phases A and B, differing in peak numbers and locations



(b) Three XRD scans for various samples of the same material. The peak position, shape and intensities vary between the variants.

Figure 2: Differences between XRD scans of distinct materials and samples of identical phases. The XRD scans serve as an input for the classifiers.

procedure, the model learns to identify the components and depending on the respective fractions, the input sample is assigned to the most similar mineral of the training set. The NMF approach, however, is developed for specific, small applications like the ternary Al-Li-Fe oxide system with 6 phases to be distinguished [2]. It is unclear whether the performance can be transferred for larger datasets, since in the field of material sciences hundreds or thousands of materials frequently have to be distinguished from each other.

Most recently, neural networks gained widespread popularity for use with image data and were also applied to XRD data. First, Park et al. [5] and Oviedo et al. [6] showed that convolutional neural networks work well for predicting crystalline properties like space group and the crystal system. Then, Wang et al. [7] demonstrated that a neural network with a VGG16 like architecture [7] is able to distinguish between 1012 metallic phases. Similarly, convolutional neural networks proved to be reliable for the prediction of unknown one-dimensional NMR signals [4].

In application, the presented networks and algorithms are able to assign an unknown sample to a known reference material. However, it is a requirement that the corresponding material is also present in the training dataset, otherwise the classifier is not able to provide a meaningful result. For the XRD case with

over 1000 learned metallic phases, there is a high probability that the unknown material was also present in the training dataset for a metallic sample, but in a typical diffraction pattern database there are more than 10000 different phases. However, it is unclear whether it is possible to train a single classifier for all possible materials, while still identifying minor differences between similar phases.

Furthermore, in the field of materials science it is common to identify new materials, which would require an extension of the existing classifiers. The presented neural networks [7, 5, 6] are not easily extensible, requiring to train the classifier again in order to add one class to the existing ones. Instead of retraining all of the layers, it is also possible to freeze the weights of the convolutional layers and only retrain the fully connected layers before the output, a method that is usually referred to as transfer learning [8].

Alternatively, a concept called Siamese Neural Networks offers the functionality to extend or limit the classes during the prediction process. Once developed for signature verification [9], Siamese neural networks feature two inputs with parallel processing operations that share its weights. Accordingly, the network rates the similarity of the two processed inputs. For a classification task, one input is the unknown sample to be analyzed and the second one a reference material. After comparing the sample to all references, the networks assigns the class by choosing the highest similarity to a known material. Consequently, the network does not require retraining to add another class, but rather an additional input to be compared is added.

The Siamese network approach is particularly popular in the field of face recognition, where the reference database is continuously supplemented with new faces [10]. Here, the Siamese network uses its convolutional layer structure to extract characteristic features from the 2D inputs to find similarities and distinguish between different faces. In the field of material science, Zhang et al. [11] used Siamese networks for application with 2D NMR data to distinguish between material classes. Two dimensional NMR data, however, is more time-consuming to measure, thus an expert is commonly used to manually evaluate the one-dimensional instead. Accordingly, we employ Siamese networks for

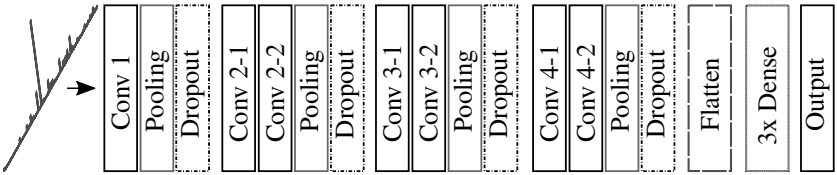


Figure 3: Architecture for VGG16-like classification network with convolutional layer (Conv) stacks, max pooling layers (Pooling) and three fully-connected layers (Dense) before the output

the first time to classify one-dimensional data describing measured materials, for which currently only classifiers exist, that are not extendable.

3 Methods

3.1 Siamese Networks

We want to evaluate whether Siamese networks are suitable for use with one-dimensional data, using XRD data as an example. To assess the performance of the Siamese network, we use the classification network developed by Wang et al. [7] as a reference, as presented in Figure 3. Here, we use diffraction patterns with 3250 datapoints (measured from 5^o to 70^o with $\Delta 2\theta$ of 0.02) as an input for the network and employ an architecture of convolutional layers, max pooling operations and dropout before we reshape the features into an one-dimensional embeddings vector in the flatten layer. For the convolutional operations, we use a 5x1 kernel with 6 to 64 filters from the first to the last convolutional layer, apply max pooling with a pool size of 2 and stride 2 and utilize dropout of value 0.2 between the pooling and convolutional layers of the different stages to reduce overfitting during the training. After the reshaping of the features into an embeddings vector, the three fully-connected (dense) layers learn the classification rules before the networks outputs the prediction scores of the classes in the output layer. Therefore, the number of neurons in the output layer corresponds to the number of eligible classes.

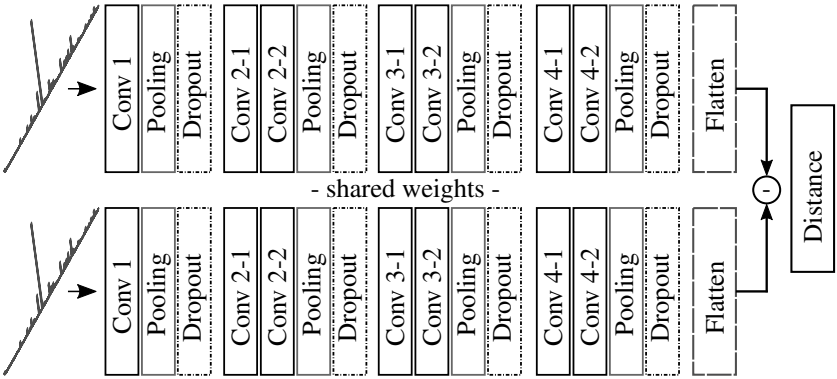


Figure 4: Architecture for VGG16-like classification network

Due to the dense layers, the classification network is not transferable to another set of classes without retraining. One concept to reduce the required retraining time is transfer learning where we freeze the weights of the convolutional layers, so they cannot be changed during the training process.

Accordingly, we try two different tactics to retrain the classification network for different classes:

1. *Full retraining* of all layers, and
2. *transfer learning* approach, where we freeze the weights of the convolutional layers and only retrain the dense layers.

Instead of learning the classification rules in the fully-connected layers, the Siamese network uses the embeddings vectors of the extracted features to rate the similarity of two fed input patterns, as visualized in Figure 4. The Siamese network employs two parallel processing pipelines to transform the input XRD scans into comparable features by sharing the weights between the corresponding layers. Thus, both inputs have to be of the same kind, the network is not able to compare a XRD scan with a crystalline properties vector.

During the training process the Siamese network has to fine-tune the weights to decrease the distance between embeddings of the same class, while increasing

the distance between classes. The distance between two vectors $\mathbf{v}$ and $\mathbf{w}$ is usually calculated using the Euclidean distance

$$d(\mathbf{v},\mathbf{w}) = ||\mathbf{v}-\mathbf{w}|| = \sqrt{\sum_{i=1}^{n}(v_i - w_i)^2}. \tag{1}$$

Thus, the network can be trained by feeding positive and negative examples for each sample and class and adapting the weights to minimize or maximize the distances accordingly. One loss function that combines this functionality into a single function to be optimized is the triplet loss [10]

$$L(A,P,N) = \max(||f(A) - f(P)||^2 - ||f(A) - f(N)||^2 + \alpha, 0). \tag{2}$$

Here, the embedding of the input sample $f(A)$ (anchor) is compared to the embeddings of a positive sample $f(P)$ of the same class and a negative one $f(N)$. The difference between the distances has to be at least equal to the margin α for the loss to be zero, otherwise it is positive. Therefore, the Siamese network is trained by minimizing the triplet loss function.

3.2 Training Data

Measured and labelled data is not available in the required amount since deep learning algorithms require several hundred training examples. Thus, we rely on synthetic XRD data, which we simulate based on a database of measured crystallite properties and physical principles, perfectly imitating measured diffraction data [12]. By simulating the XRD scans, we ensure that the training data contains the relevant deviations of diffraction peaks demonstrated in Figure 2b and every class is represented with the same number of scans. Accordingly, we assemble two datasets:

For the first dataset A, we select 100 random crystallite materials from a reference database (which contains 345 materials in total) and simulate 50 variations for each, resulting in 5000 total synthetic XRD scans. The training, validation and test set are made up with a 50-20-30 split, so for every class 25

variations are used for the training set, 10 for the validation set and 15 for the test set. In total, dataset A contains 2500 samples in the training set and 1000 and 1500 samples in the validation and test set respectively. Accordingly, we ensure that no class is under-represented in any of the sets.

Secondly, we set up dataset B in the same way. We select another 100 random crystallite materials without any overlap between the classes of datasets A and B. Again, we split the 5000 synthetic XRD scans into training, validation and test data using the 50-20-30 split.

Using datasets A and B, we evaluate the performance of a classifier and Siamese networks, as well as transfer learning and the transferability of the Siamese network. Hence, we train the classifier and Siamese network with the train set of dataset A and choose the weights that perform best on the validation set to avoid overfitting. Afterwards, we assess the performance of the VGG16-like classifier and the likewise Siamese network for dataset A using the test set. For the Siamese network, we pick a random scan of the train or validation set as the reference input for each class to calculate the distances for the test samples. As a performance metric, we choose the Top-1 and Top-3 accuracy for the highest prediction scores of the classifiers and the smallest distances for the Siamese network.

Subsequently, we apply the trained Siamese network for the test set of dataset B without retraining. Before we evaluate the performance of the classifier for dataset B, we retrain the network using the two previously described tactics to evaluate the transfer learning approach. Finally, we rate the ability of the Siamese networks for transfer between datasets in comparison to the classifiers performance with full-retraining or transfer learning as an alternative.

4 Results

After training the classification and Siamese network using the train and validation XRD scans, we evaluate the Top-1 and Top-3 accuracy for the 1500 diffraction patterns of the test set. Table 1 presents the performance scores for both networks. Overall, we achieve near perfect results for the classifier with

Table 1: Comparison of Top-1 and Top-3 performance between the VGG16-like classifier and Siamese network for dataset A.

	Network Type	
Accuracy	Classifier	Siamese
Top-1	99.9%	97%
Top-3	100%	99.9%

a 99.9% Top-1 accuracy and can be sure that the correct phase is within the three highest predicted classes (100% Top-3 accuracy). The performance of the Siamese network is slightly worse than the classifier, with scores of 97% and 99.9% Top-1 and Top-3 accuracy, respectively. The results of the Siamese network suggest that there are materials with very similar diffraction patterns in the dataset, which the Siamese network can hardly distinguish. Accordingly, we obtain an almost perfect Top-3 accuracy, because the network does not have to decide correctly. In contrast, the classification network is able to distinguish the subtle differences by means of the dense layers, which means that there is hardly any difference between Top-1 and Top-3 accuracy.

Next we evaluate dataset B, for which we can use the Siamese network without having to retrain. The classification network is retrained using the full retraining and transfer learning approach, where we freeze the weights of the convolutional layers during the training process and only fine-tune the dense layers. Table 2 shows the prediction scores for the two retraining strategies of the classifier and the Siamese network without retraining. For the full retraining approach of the classifier, we achieve almost identical accuracies compared to dataset A, with 98.7% and 99.9% respectively. The slightly worse Top-1 accuracy indicates that the phases of dataset B are harder to distinguish from each other. Notably, the Siamese network achieves an almost identical performance for dataset B with a 99.3% Top-3 accuracy, although it was trained with phases of dataset A only.

However, the transfer learning strategy is only partially suitable in the case of transferring XRD datasets. Although the classifier still manages to predict the correct phase in about 70% of all cases, it is significantly worse than the full retraining approach. This leads to the conclusion that the weights of the

Table 2: Comparison of Top-1 and Top-3 performance between the VGG16-like classifier and Siamese network for dataset B.

	Network Type		
	Classifier		Siamese
Accuracy	Full Retraining	Transfer Learning	
Top-1	98.7%	70.4%	93.5%
Top-3	99.9%	70.7%	99.3%

convolutional layers were strongly specialized to emphasize the peculiarities of dataset A. Consequently, the network is no longer able to highlight the differences in the phases of dataset B in the transfer learning approach, so that the dense layers can use them for the classification.

Furthermore, the classifier requires more time to be applied to dataset B, regardless of the retraining strategy. Firstly, 2500 synthetic training scans (+ 1500 validation scans to avoid overfitting) must be generated so that the network learns the classification rules for the materials of dataset B. In comparison, the Siamese network requires only one reference per material, so 100 synthetic scans. Secondly, the retraining process for the classifier takes some time[1], while the Siamese network is applied to the second dataset without retraining.

Overall, the Siamese networks prove that they are well suited for the use with XRD data and can also be used for materials that were not represented in the training dataset. Since the diffraction patterns of identical materials can differ significantly, as shown in Figure 2b, it is of interest to know how the Siamese network processes the XRD scans to achieve the small distance between the embedding vectors of identical materials. Hence, we visualize two embedding vectors for XRD scans of the same material that differ in peak position, shape and intensity in Figure 5. Interestingly, the Euclidean distance between the two

[1] The duration of the training process is strongly dependent on the utilized computing power. Since the Siamese networks do not require to be retrained, we refrain from comparing the absolute training times.

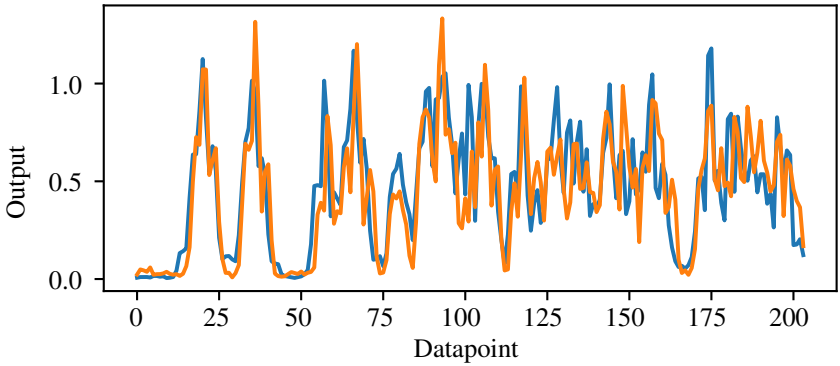


Figure 5: Embedding vector of the Siamese network for two XRD scans of the identical material.

vectors is definitely not zero, as there are clear differences between the network outputs. The output of the network reminds of the input XRD scans, which, however, are shrunk in spatial resolution by the Max Pooling layers. This presumably enables the network to compensate for the possible shifts of the diffraction angles, although the calculated peaks are not completely aligned.

5 Conclusion and Outlook

In summary, we employ Siamese neural networks for the first time with one-dimensional XRD data to distinguish between materials. Therefore we trained a reference classification network on two synthetic datasets and show that the Siamese Network achieves almost identical results. The strength of the Siamese network lies in the fact that it can be applied to materials not included in the training dataset without the need for retraining of the network (in comparison to the classification network). Accordingly, we demonstrate that the Siamese networks are particularly suitable for use in the field of material sciences, where the classes are dynamically composed. The comparable transfer learning approach has not proven to be adequate in our case because the features learned in the classifier were too specific for transferability.

Lastly, we showed how the neural network computes the embedding vectors, and that the Siamese network is not yet able to fully compensate for the dif-

ferences within diffraction patterns of the same material. Therefore, it has to be investigated whether the distance calculation using the pairwise, Euclidean distance for the one-dimensional data can be replaced by a more suitable alternative. In the field of time series analysis, for example, there is a distance calculation using so-called Dynamic Time Warping, which compensates for small displacements on the time axis. It is possible that such a function enables the network to compensate for the differences even better and thus differentiates the materials even more clearly.

The next step is to validate our results with measured XRD scans. While it is necessary that as many real possible variations as possible are represented in the training data and hence, synthetically generated data are possibly essential for the training of a Siamese network, the trained deep learning model should be able to use measured XRD scans as a reference input.

Moreover, we have trained Siamese networks using XRD data as an example, but a deployment with other one-dimensional data would also be imaginable. Especially for similar analytical methods like NMR, for which a classifier based on a neural network has already been developed, a transfer of our results would be conceivable. Accordingly, the next step is to test our Siamese network approach with other one-dimensional data.

References

[1] V. Pecharesky and P. Zavalij. *Fundamentals of Powder Diffraction and Structural Characterization of Materials*. Springer Science & Business Media, 2005.

[2] V. Stanev, V. Vesselinov, A. Kusne, et al. Unsupervised phase mapping of x-ray diffraction data by nonnegative matrix factorization integrated with custom clustering. *npj Computational Materials*, 4:1–10, 2018.

[3] H. Wang, Y. Xie, D. Li, et al. Rapid identification of x-ray diffraction patterns based on very limited data by interpretable convolutional neural networks. *Journal of Chemical Information and Modeling*, 60:2004–2011, 2020.

[4] Bruker Biospin Corporation. Deep learning applications in nmr spectroscopy. In *European Magnetic Resonance Meeting*, 2019.

[5] W. Park, J. Chung, J. Jung, et al. Classification of crystal structure using a convolutional neural network. *IUCrJ*, 4:486–494, 2017.

[6] F. Oviedo, Z. Ren, S. Sun, et al. Fast and interpretable classification of small x-ray diffraction datasets using data augmentation and deep neural networks. *npj Computational Materials*, 5:1–9, 2018.

[7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR*, 2015.

[8] S. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[9] J. Bromley, I. Guyon, Y. LeCun, et al. Signature verification using a Siamese time delay neural network. In *Advances in neural information processing systems*, 1994.

[10] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[11] C. Zhang, Y. Idelbayev, N. Roberts, et al. Small molecule accurate recognition technology (SMART) to enhance natural products research. *Scientific Reports*, 7(1):1–17, 2017.

[12] J. Schuetzke. Evaluation of machine learning approaches for crystalline phase identification. *Master's Thesis, Karlsruhe Institute of Technology*, 2019.