Siamese Neural Network Improves the Performance of a Convolutional Neural Network in Colloidal Self-Assembly State Classification

A. Lizano-Villalobos,[1, a)] B. Namikas,[2, a)] and X. Tang[1]
[1)] *Cain Department of Chemical Engineering, Louisiana State University, Baton Rouge, LA, 70803 USA*
[2)] *Baton Rouge Magnet High School, Baton Rouge, LA, 70806*

(*Electronic mail: xuntang@lsu.edu.)

(Dated: 11 November 2024)

Identifying the state of the colloidal self-assembly process is critical to monitoring and controlling the system into desired configurations. Recent application of convolutional neural networks with unsupervised clustering has shown comparable performance to conventional approaches, in representing and classifying the states of a simulated 2D colloidal batch assembly system. Despite the early success, capturing the subtle differences among similar configurations still presents a challenge. To address this issue, we leverage a Siamese Neural Network to improve the accuracy of the state classification. Results from a Brownian dynamics-simulated electric field-mediated colloidal self-assembly system, and a magnetic field-mediated colloidal self-assembly system demonstrate significant improvement from the original convolutional neural network-based approach. We anticipate the proposed improvement to further pave the way towards automated monitoring and control of colloidal self-assembly processes in real time and real space.

I. INTRODUCTION

Self-assembly of nano- and micron-sized particles into macroscopic systems has potential applications in renewable energies, health care, advanced computing, and biotechnology.[1–5] The property of the self-assembled materials arises from both the property and the configuration of their constituent particles.[6–8] While manipulating particle movement, thus the assembly configuration, is feasible with external forces, such as electric, magnetic, and acoustic fields, assembling nano- and/or micron-sized particles into specific configurations is still challenging. A first step towards this task is to have accurate description of the assembly structure, which is also commonly referred to as the state representation.

Conventional approaches for state representation of colloidal self-assembly include physics-based order parameters,[9–13] such as C_6 and Ψ_6,[14,15] which are mathematical equations based on the particle coordinates, and metrics derived from mathematical and statistical techniques such as Principal Component Analysis,[16–20] UMAP,[21–24] or other dimensionality reduction techniques,[25–29] that normally bear no physical meaning. Recent methods that combine partial physical understanding with machine learning algorithms to learn patterns, from data generated either from simulations or experiments have shown promise and have started to attract attentions in the field.[30–34] Examples include neighborhood graphs,[35–37] spherical harmonics,[38–40] and many other descriptors. These descriptors are normally obtained from the spatial coordinates of the particle,[41,42] and are further processed with dimensionality reduction and clustering techniques for a low dimensional representation of the colloidal assembly state, with physically explainable meanings.[30,33,34] While these approaches have shown success in representing the colloidal self-assembly state, they require the identification of particle coordinates and relevant physical descriptors, which is challenging in large systems or systems with smaller particles and complex dynamics.

On the other hand, images provide information about the configuration without knowing the specific location of every constituent particle, and image-based structure analysis has started to gain popularity. Recent examples include using images to estimate the density and size of assembled structures for lithographed colloidal cones,[43] and to predict the temperature of isothermal heating processes on polymer colloidal crystals.[44]

In Ref. 45, we introduced a Convolutional Neural Network (CNN)-based framework that combines image analysis with unsupervised clustering, to classify different colloidal self-assembly configurations, and showed on par performances to the use of order parameters for a Brownian dynamics simulated 2-dimensional electric field-mediated system.[45] In that framework, images are first projected into a set of low-dimensional features through a Convolutional Autoencoder (CAE), then these low-dimensional features are clustered with HDBSCAN[46] to group similar images to form a set of clusters/states the system could assemble into. Within each cluster, the image with the smallest average distance to all other images is named as the centroid image, representing the typical configuration of that cluster. Each image is then labeled according to its cluster label before being used to train a CNN classifier, which would classify a newly obtained image into one of the defined cluster/state.[45]

While the CNN-based framework proposed in Ref. 45 showed success in distinguishing states that can be challenging to distinguish with order parameters, e.g. void defects vs. ordered crystals, it still struggled to provide 1) a high-purity clustering, where each cluster would only contain one type of configurations, and 2) a non-redundant clustering, where no multiple clusters would contain the same type of configurations. While the redundant clusters could be easily resolved with visual inspection after the clustering (assuming the total number of the resulting clusters is manageable), improving the

[a)] Co-first authors

purity of each cluster is nontrivial, as it requires capturing the subtle differences at the local features of the configuration. To address this issue, here we introduce a modified CNN-based framework which features the use of a Siamese Neural Network (SiNN) to detect the subtleties of similar configurations for improved purity in the resulting clusters.

A SiNN is a model that maps images into a low-dimensional representation, while preserving the similarities between the local features of the images.[47] SiNN models have gained popularity in applications such as the detection of laser induced defects in optical elements,[44,48] the identification of glass transition on liquid crystal systems[49] and many others.[47,50] The SiNN learns similarities from groups of three images called triplets, which consist of an anchor, a positive and a negative image. The anchor serves as a reference image, the positive image is known to be similar to the anchor, and the negative image is different from the anchor and the positive. The SiNN is trained by minimizing the distance between the anchor and the positive image, while maximizing the distance between the anchor and negative image at the same time. In this study, we introduce a SiNN as an intermediate step after the unsupervised HDBSCAN clustering and before the training of the CNN model, as depicted in Fig. 1

Specifically, images from the unsupervised clusters will form the triplets used to train the SiNN, to re-assign the mislabeled images to their correct clusters, for an improved classification. Validation on both a Brownian dynamics simulation model and two sets of experimental images demonstrated, that the updated classification framework with the SiNN significantly improved the accuracy of the state classification, as compared to the one without SiNN. We anticipate the proposed framework to further benefit systems with more complex dynamics and states.

II. DATA GENERATION AND PERFORMANCE EVALUATION METRICS

The framework was developed with data from a Brownian Dynamics (BD) simulation model, same as in Ref. 15 and Ref. 45, which simulates the 2-dimensional (2D) assembly dynamics of SiO_2 colloidal particles in a quadrupole electric field-mediated batch container. A total of 13765 images were generated from the simulation, with 90% of the total images used as the training set to develop the neural network models, and the remaining 10% used for validation. The performance of the framework is also tested with two sets of experimental data. The experimental datasets were collected from the same system as in Ref. 51, which features a combination of an electric and a magnetic field. In this system, a sinusoidal AC electric field is applied to the x-y plane, with an AC function generator connected between the top and bottom electrodes around the container. The magnetic field is generated through four air-cored copper solenoid coils of 50 mm inner diameter, 51 mm length, and 400 turns with a current capacity of 3.5 A, installed into a polyacrylic crossing board. Each coil is connected to a DC or AC function generator, which is controlled by a custom Python code with an output card. The magnetic field strength was measured by a Gauss meter. To generate the circularly rotating magnetic field, 80 Hz AC currents are applied to four coils. The phase lag between neighbouring coils is ±90° so that both clockwise and counterclockwise rotational magnetic fields can be generated. The detailed description of the system and dynamics can be found in Ref. 51. In this system, thousands of superparamagnetic particles with a diameter of $3\mu m$ are exposed to a planar rotating magnetic field, where the field is increased from 1 to 3 mT to permit assembly from dispersed states to the conjugated configurations. The first experimental dataset contains 2945 images, and the second dataset contains 2291 images, collected from the same system but under a different field strength. Fig. 2 gives the typical configurations observed in these two systems.

To gauge the clustering performance, we introduce the concept of purity to quantify the percentage of images that are correctly clustered, which is defined as: $Purity_i = \frac{N}{M}$, where M is the total number of images in cluster i, and $N \leq M$, is the number of images that are correctly classified to cluster i. In our previous study,[45] $Purity_i$ was calculated by comparing the images within cluster i to the centroid image of cluster i. While this definition provided a measure of purity within each cluster, it has the potential to overlook intra-cluster misclassification. To avoid this, here we manually pre-labelled all images in our dataset to one of the pre-identified representative configuration clusters as in Fig. 2. Such a pre-labeling requires considerable human efforts, but it could provide a more unbiased evaluation of the clustering performance. Note that although nine different typical configurations are identified in Fig. 2a, the unsupervised clustering might not yield exactly nine clusters, and some of the configurations could be grouped into one, depending on the resolution of the classification. For example, Cluster 1 and 2 could be grouped together, Cluster 3 and 4 could be merged together, and Cluster 5, 6, and 7 could also be grouped into a single cluster. Furthermore, the configurations given in Fig. 2a only show commonly encountered configurations, and do not provide an exhaustive coverage over all different configurations. For example, it is possible that multiple-void defective states exist besides the single and double-void defective states as shown in Fig. 2a. To deal with scenarios where the unsupervised clustering yields a different number of clusters, we compare the centroid images of the defined clusters against the nine pre-defined clusters to re-assign the appropriate labeling, based off the initial pre-labeling, and the resulting re-assigned labels are used as the ground truth labeling of each image, in calculating the ultimate purity of each cluster. We demonstrate the details of this process with the BD-simulated colloidal self-assembly state classification.

III. SEMI-SUPERVISED TRAINING OF SINN VIA TRIPLET MINING

The SiNN model consists of three convolutional layers, each with a ReLu activation and a 2x2 MaxPooling layer. We used 32, 64 and 128 channels for each convolutional layer, and a 3x3 kernel size for all of them. The embedding of the model, based on which the clustering was performed, is the

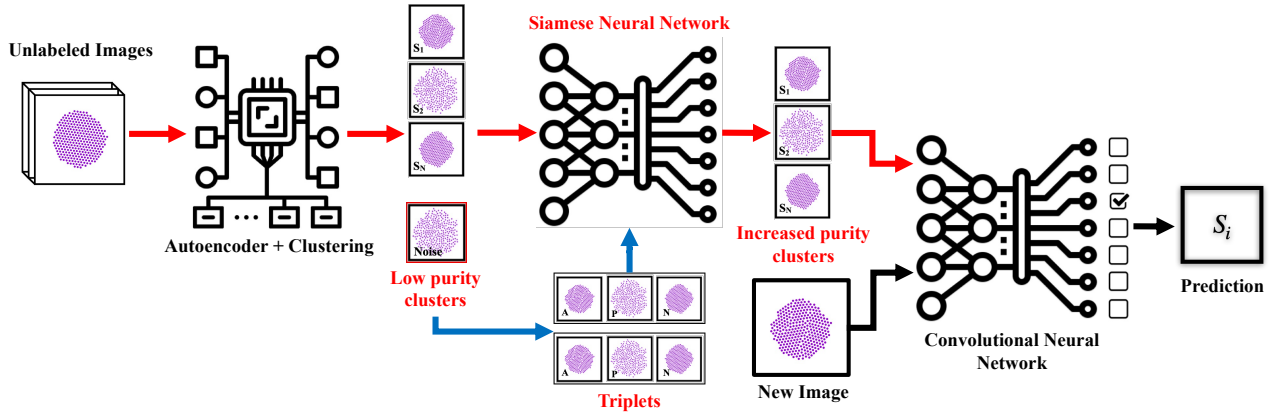

FIG. 1. The proposed modification includes a SiNN model (labeled in red) trained by minimizing the triplet loss function. The triplets used to train the model are generated based on the clusters obtained from clustering the autoencoder features (blue path). The labels used to train the CNN model are obtained from the relabeling process done with the SiNN model.

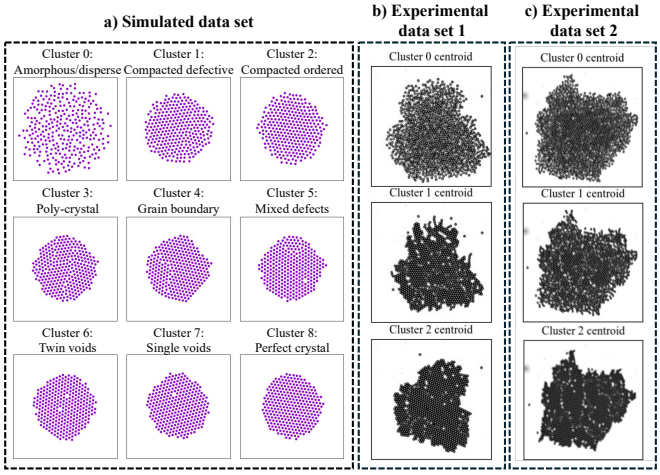

FIG. 2. Representative configurations from the dataset for pre-labeling the images. (a) Nine states are identified for the BD-simulated electric-field mediated assembly system. (b and c) Three states are identified for the planar rotating magnetic field-mediated assembly system.

output of a densely connected layer with 64 neurons and a linear activation. The model parameters are updated by minimizing the triplet loss function defined in Eqn. 1, using the Adam optimizer.

$$\mathscr{L}(Z^+,Z^\pm,Z^-) = max\{d(Z_n^\pm,Z_n^+) - d(Z_n^\pm,Z_n^-) + m, 0\} \quad (1)$$

In the triplet loss function $\mathscr{L}(Z^+,Z^\pm,Z^-)$ in Eqn. 1, images are embedded in vectors Z, where "+" indicates the positive image, "±" indicates the anchor image, and "−" indicates the negative image. $d(Z_n^\pm,Z_n^+)$ is the difference between the anchor and the positive image, and $d(Z_n^\pm,Z_n^-)$ is the difference between the anchor and the negative image, in triplet set n, out of the N total sets of triplets. The loss is then calculated as the maximum between 0, and the distances augmented by a margin $m = 3.0$, which facilitates learning while preventing overfitting.

To generate the N sets of triplets, we performed HDBSCAN clustering, based on features extracted with an autoencoder, as described in our previous work.[45] The anchor image for each triplet is the centroid image of each cluster, and the positive image is taken from the proximity of the centroid image, whereas the negative image is taken from a different cluster. For example, given a set of K clusters, and by taking L sample images for the positive and negative images, we obtain $K(K-1)L^2$ triplets. To reduce the impact of rotational changes in the images, we also included randomly rotated positive images in the training sets. For the current work we trained our model using the seven clusters obtained in our previous work[45] (i.e. $K = 7$) and selected 10 samples per cluster (i.e. $L = 10$), to form a data set of 4200 triplets in total.

IV. RESULTS

A. Validation with the BD simulated 2D electric field-mediated system.

We first investigated the performance of modified framework with the BD simulations for an electric field-mediated 300 particle colloidal self-assembly. To perform HDBSCAN clustering without the SiNN, we used the autoencoder approach from Ref. 45 to extract features from the configuration. Fig. 3a shows the centroids of each cluster along with the pre-labels that the centroid could be classified as, and the histograms in Fig. 3b show the pre-label distribution of the images in each cluster. These clusters were then processed with the SiNN to generate the new sets of clusters, whose centroids are shown in Fig. 3c, and the corresponding pre-label distribution histogram is shown in Fig. 3d. Comparing the histogram distribution, we observed evident improvement in the purity for all clusters, noticing the reduced number of bars in Fig. 3d as compared to Fig. 3b, which indicates the

TABLE I. Purity distribution for each cluster obtained with the autoencoder features/HDBSCAN and after SiNN relabeling

Autoencoder				SiNN			
Cluster	**Size**	**Hits**	**Purity**	**Cluster**	**Size**	**Hits**	**Purity**
0	537	534	99.4%	**0**	897	890	99.2%
1	82	28	34.1%	**1**	21	21	100.0%
2	168	35	20.8%	**2**	22	17	77.3%
3	85	25	29.4%	**3**	46	38	82.6%
4	166	126	75.9%	**4**	125	125	100.0%
5	119	40	33.6%	**5**	124	58	46.8%
6	220	120	54.5%	**6**	142	123	86.6%
Overall	**1377**	**908**	**65.9%**	**Overall**	**1377**	**1272**	**92.2%**

elimination of misclassified images.

To quantify the purity, we first relabeled all the images as the following. For the framework without SiNN, images with a pre-label of 0, 1, or 2 are all relabeled as Cluster 0, to reflect the fact that only one cluster from the unsupervised clustering has a centroid that represents the fluid-like structure (Cluster 0 in Fig. 3a). Images with a pre-label of 5, 6, and 7 are relabeled as 1 if they are in Cluster 1, or 2 if in Cluster 2, or 5 if in Cluster 5, as these three clusters are all for void-defect images, etc. The same relabeling was then conducted for the SiNN-based clusters, and with this re-processing, we then calculated the purity for each cluster.

Table I compares the purity of all the clusters with and without SiNN. With the SiNN further processing, we achieved an overall purity (defined as the percentage of the correctly classified images over the total number of images in the entire dataset) of 92.2%, showing a relative percentage improvement of 40% (defined as (92.2%-65.9%)/65.9%) as compared to that without SiNN. All clusters showed an improvement with the SiNN, especially in Clusters 1, 2, 3, and 6, each with a relative percentage improvement of 193%, 271%, 181% and 59%, illustrating the effectiveness of using SiNN to improve the classification accuracy. Despite obvious improvement, we also noticed a low purity in Cluster 5. Visual inspection of images in this cluster reveals a mix of void-defect configurations, perfect crystals and grain boundaries. To understand this behavior, we projected the autoencoder features into a two-dimensional space using t-SNE dimensionality reduction. The projection, shown in SI Fig. 1, illustrates how areas near the boundary regions between clusters overlap. These overlapping areas often coincide with transitions between states, such as amorphous to polycrystalline, grain boundary to void defective, or perfect crystal to void or grain defect, and still presents a challenge for SiNN to distinguish.

After verifying the improvement in clustering, we then proceed to classifying new images. Same as in our previous work,[45] a new incoming image is labeled with a CNN model trained on the clusters. The output of the CNN is a set of unnormalized scores, called logits, which are related to the likelihood of an image to represent a certain state. The incoming image is labeled to the state with the highest logit value, as defined in Eqn. 2, where S_i is the state assigned to image i and l_i^c is the c-th logit for image i.

$$S_i = \arg\max_c (l_i^c) \tag{2}$$

The performance of the SiNN-modified CNN-framework was evaluated with respect to two aspects. First, we investigated the prediction accuracy of the SiNN against labels obtained from the SiNN clustering (Fig. 4.a), to see how the SiNN-based CNN model was trained. Second, we evaluate the prediction accuracy of the SiNN against the ground truth labeling of each image, i.e. the re-assigned labels (Fig. 4.b).

The confusion matrices in Fig. 4 show an excellent overall prediction accuracy of 99.4%, using the SiNN labels, and an overall prediction accuracy of 92.1%, using the re-assigned True labels. Similar to what we observed in the purity analysis, the model yielded high accuracy for clusters with a high purity and lower accuracy for clusters with a lower purity, such as Cluster 5. This phenomenon suggests the carryover of the clustering inaccuracy to the prediction, and highlights the necessity of having a high purity clustering in the framework.

B. Validation with orthogonal electric and magnetic fields-mediated paramagnetic colloidal self-assembly

As the approach presented in this work analyzes static images of the assembly to identify the system state, differences in the underlying dynamics of the system would therefore not hinder the application of the framework to a new (where the assembly dynamics are governed by different mechanisms) but related system (where the assembled configurations do not differ dramatically). To deal with the new configurations arise in the electric-magnetic combined field-mediated assembly process, we adopt transfer learning to adapt our model for this new system. Transfer learning is a machine learning technique where a model developed for a particular task is reused as the starting point for a model on a different task. Essentially, it involves taking knowledge learned from one problem to apply to a different but related problem.[52–54] Here we leveraged this capability of transfer learning to use the Neural Networks trained over the electric field-mediated system as the starting point for the electric-magnetic field-combined experimental system described in II. Both datasets were obtained from the same experimental setup, differing only in the number of particles in the system and the field strength. Similar to the BD-simulated images, to enable evaluation of the performance, we conducted visual inspection of these datasets to obtain pre-labels for each image. With this, we then identified three representative configurations for each dataset (Fig. 2b and c), showing the evolution of the system from the amorphous state (Cluster 0) to the defective states (Cluster 2).

Considering the limited amount of experimental images, we split each dataset to 30%-70% for the training and validation purpose, and applied transfer learning to facilitate the model development. Specifically, we first used transfer learning to train an autoencoder with the experimental images, based on the one trained for the BD-simulated system. We then used

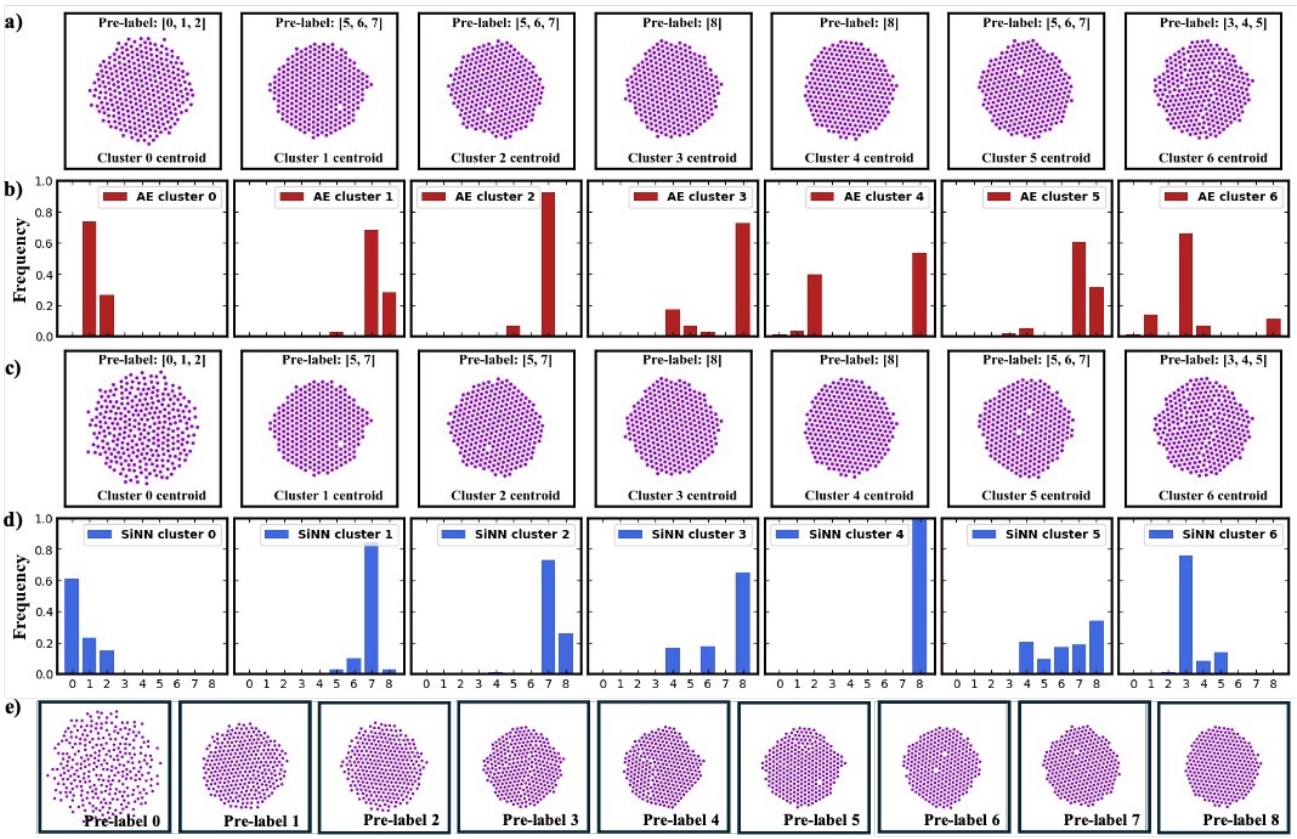

FIG. 3. Ground truth label distribution for clusters with and without SiNN relabeling. (a) Centroid images for the clusters obtained from HDBSCAN clustering without SiNN relabeling. (b) Histogram distribution of the pre-labels of clusters obtained with HDBSCAN clustering. (c) Centroid images for the clusters obtained with SiNN relabeling. (d) Histogram distribution of the pre-labels of clusters obtained with SiNN relabeling. (e) Typical configurations for each of the nine pre-identified clusters.

features extracted from the autoencoder to perform clustering with HDBSCAN, and obtained six (Fig. 5a) and four (Fig. 5.b) clusters for the two datasets, respectively. With these clusters, a SiNN was then trained to further improve the purity. Note that, as the differences in the assembly from the original and the new systems signify, for example, particle shape and types of defects, more images from the new system would be needed to retrain the model with transfer learning for a satisfactory performance.

As we obtained more clusters than the pre-identified clusters in Fig. 2, to quantify the purity, we conducted relabeling of all the images, same as for the BD-simulated system. Specifically, inspection at the centroid of each cluster in experimental dataset 1 suggests Cluster 0,1,2,4, and 5 could effectively be treated as one cluster. Therefore, we relabeled images with a pre-label of 0 to Cluster 3, and images with a pre-label of 1 or 2 to 0 or 1 or 2, or 4 or 5, depends on which of the five clusters they locate. For the experimental data set 2, we relabeled images with a pre-label of 0 to Cluster 3, images with a pre-label of 1 to Cluster 2, and images with a pre-label of 2 to either 0 or 1 depending on which cluster they locate. The purity statistics in Table II demonstrated a marginal improvement with the SiNN over that without SiNN, as the purity with HDBSCAN clustering is already approaching 100% for all clusters. This high accuracy for the HDBSCAN clusters may be due to the simpler dynamics observed in the samples collected here, as compared to that from the BD-simulated system.

To develop the CNN model for new image classification, we again used transfer learning where we only modified the output layer of the CNN model developed for the BD-simulated system, as the number of clusters for each data set differs. The CNN was then retrained by reducing the cross-entropy loss between the SiNN labels and the logits predicted from the CNN model, with the training experimental dataset. The confusion matrices in Fig. 5 show an excellent prediction accuracy on the two experimental data sets, with 100% for experimental data set 1 and 99.4% for experimental data set 2. Admittedly, the relatively simple configurations in the data sets would have contributed to this near-perfect accuracy, such a high prediction accuracy still demonstrates the reliability of the framework presented here.

V. CONCLUSION AND DISCUSSION

In this study, we demonstrated the efficacy of a modified image-based state classification method for colloidal self-

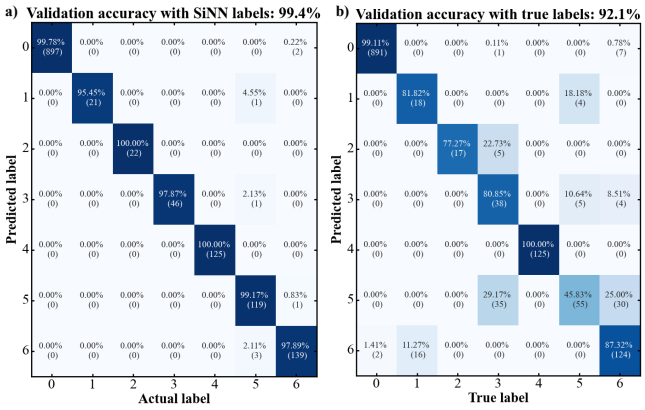

FIG. 4. Classification of validation data set using the CNN model. a) Confusion matrix showing how the labels assigned by the CNN (predicted labels) compared with the labels assigned by the SiNN (actual labels). b) Confusion matrix showing how the labels assigned by the CNN correlate with the labels from the ground truth labeling (true labels).

TABLE II. Purity distribution for each cluster obtained with and without SiNN relabeling.

Experimental Data Set 1

Autoencoder				SiNN			
Cluster	**Size**	**Hits**	**Purity**	**Cluster**	**Size**	**Hits**	**Purity**
0	75	75	100.00%	0	311	311	100.00%
1	92	86	93.48%	1	1108	1108	100.00%
2	60	60	100.00%	2	127	127	100.00%
3	62	62	100.00%	3	94	94	100.00%
4	1668	1642	98.44%	4	173	173	100.00%
5	90	90	100.00%	5	248	248	100.00%
Overall	**2061**	**2029**	**98.45%**	**Overall**	**2061**	**2061**	**100.00%**

Experimental Data Set 2

Autoencoder				SiNN			
Cluster	**Size**	**Hits**	**Purity**	**Cluster**	**Size**	**Hits**	**Purity**
0	207	207	100.00%	0	417	417	100.00%
1	990	990	100.00%	1	945	945	100.00%
2	343	332	96.79%	2	172	171	99.42%
3	63	63	100.00%	3	69	69	100.00%
Overall	**1603**	**1592**	**99.31%**	**Overall**	**1603**	**1602**	**99.94%**

assembly systems. Specifically, the framework consists of unsupervised clustering with HDBSCAN for initial clustering, using the image pixels as the input, and a Siamese Neural Network to further improve the purity of each cluster. A convolutional neural network is then trained with the processed clusters to perform new image/state classification. Implementation of this approach on the Brownian dynamics simulated electric field-mediated system and the experimental datasets from a magnetic field-mediated assembly system, reveal that the SiNN can capture subtle differences missed by the HDBSCAN, thus improving the accuracy of the state classification. Modifying the framework developed for the Brownian dynamics simulation with transfer learning for applications on the experimental datasets, demonstrated the applicability of the approach to similar yet different systems, as well as its adaptability with amiable efforts.

Despite the improved performance, we still noticed persistent misclassifications with the SiNN, indicating the need for a better set of features. While we anticipate including more images to retrain the models could lead to better accuracy, inspecting the logits from the CNN model suggested that using multiple logits instead of the biggest logit value might be an alternative for better classification. Fig. 6 shows two misclassified configurations and the associated logit distribution from the SiNN for the BD-simulated system, i.e. results presented in Fig. 3c and d. Both configurations in Fig. 6 should be predicted as Cluster 6 (polycrystalline) in Fig. 3c, which has the second highest logit value in the distribution (circled in red). However, the use of the biggest logit value (circled in green) has misclassified them into an amorphous state (Cluster 0) and a double void defect state (Cluster 5), respectively. These observations suggest that single logit value might not be sufficient to make an accurate classification whereas including more logit values might provide a more informed classification.

Redundant clusters are another issue persistently observed in the current and previous CNN-based framework.[45] While it is possible to reduce the number of redundant clusters by manipulating the hyperparameters of the HDBSCAN, it could also jeopardize the classification accuracy at the same time. Inspecting the centroid image of each cluster and manually merge clusters with similar centroids provides an alternative, however challenge arises when inspector-to-inspector bias is nonnegligible, and the total number of clusters is huge that visual inspection becomes time-consuming.

While we expect variations over the particle properties, such as particle shape and polydispersity etc., as well as increased complexity in the system dynamics, such as the combined use of multiple fields, to not limit the applicability of our approach, as it analyzes images to perform state classification. Theoretically, as long as the training set for the machine learning models (especially for the training of the Siamese Neural Network) includes a sufficiently large amount of images that cover the most important and commonly encountered configurations in the system, the performance of the approach should be preserved. However, increased complexity in the system dynamics and variations in particle properties could potentially lead to more complicated configurations, which would require more training images and efforts to ensure the accuracy of the state classification. Furthermore, training machine learning models over a large set of images could also pose a computational challenge in terms of computing speed, storage, and cost. Transfer learning could be a potential solution, as demonstrated in this work. However, further validation is needed with more complex systems, and efforts towards improving the computation efficiency would also tremendously benefit the application of the approach to advanced systems. We also note that in the simulated electric field-mediated assembly system, we observed issues with handling the similar type of configurations but with different rotations. For example, the framework failed to correctly classify some of the same configuration but rotated to dramatically different angles (therefore, the none-perfect purities in the clusters). While we

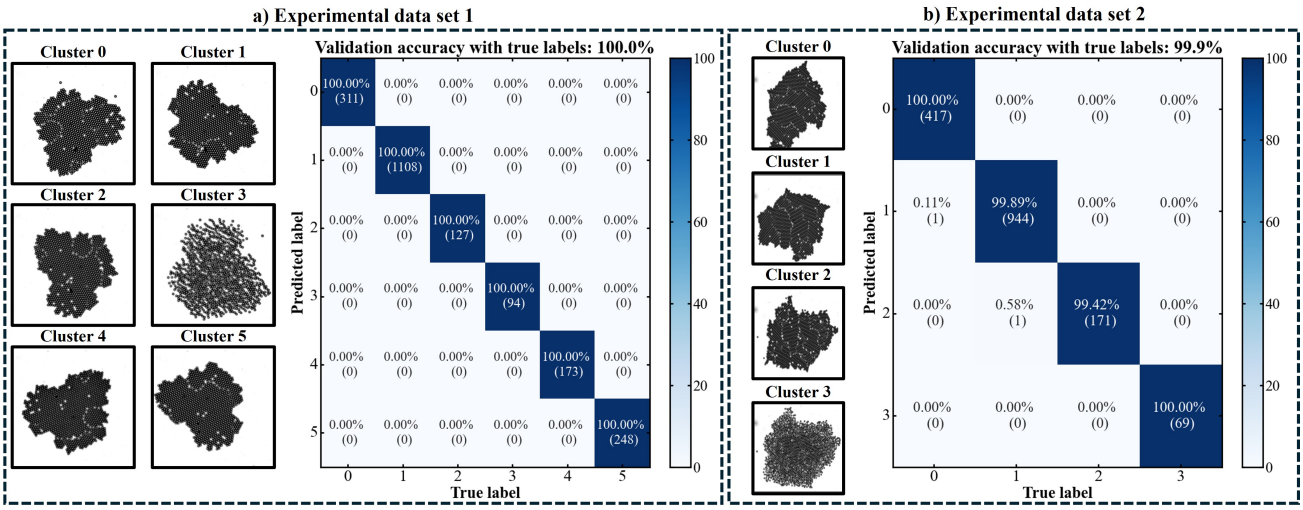

FIG. 5. Classification of validation data set using the CNN model, for both experimental data sets 1 and 2, demonstrating a high prediction accuracy. The configuration images show the centroid configurations of each of the identified cluster.

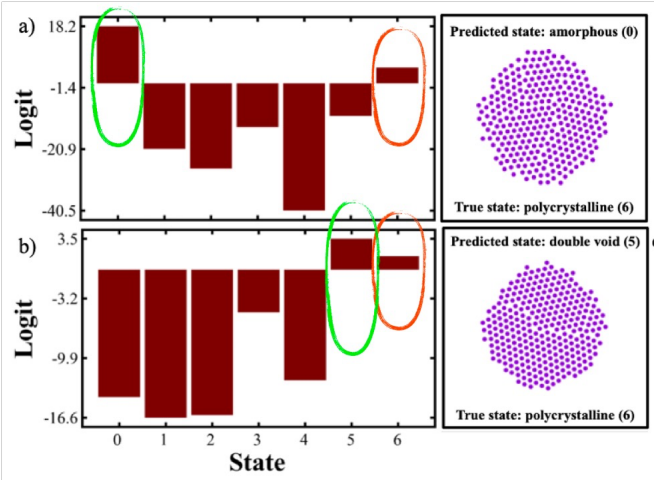

FIG. 6. Logit value distribution of two misclassified configurations from the BD-simulated colloidal self-assembly system, suggesting a potentially better feature than the single biggest logit for the classification.

expect a targeted training of the machine learning models with the rotated images to improve the performance, to account for all different rotations, we could again face a dramatic increase in the number of images, thus a computational challenge. Further investigations on how to handle such rotational variations would be important to improve the performance of the proposed approach. Lastly, as our approach relies on images for the classification, the quality of images could significantly affect the performance. While we expect images with decent quality to be attainable with current microscopy technology, we do anticipate images with compromised quality from a real system. For these scenarios, we would expect image processing to benefit the performance. For example, uneven illumination could be handled with histogram equalization and homomorphic filtering,[55,56] and the optical aberrations could be addressed with wavelet transforms and deconvolution.[57,58] Furthermore, machine learning techniques have also shown success with improving image quality, such as the use of a denoising autoencoder for noise filtration.[59] Given the modularity of our framework, addition of an image processing step can be conveniently achieved, however, validation with real process images would further benefit the development of the proposed approach with enhanced confidence.

Using unsupervised learning to perform many-body self-assembly state classification is in its infancy, and faces many challenges including methods of validation, feature selection, and accuracy improvement. However, we anticipate the potential benefits of having data-driven order parameter-free state classification approaches, such as automation of the state classification, transferability to different systems etc., to attract future efforts in this direction.

VI. SUPPLEMENTARY MATERIAL

In the Supplementary Material, we provide a two-dimensional t-SNE projection of the image features used for the clustering, demonstrating the overlapping phenomenon at the boundary of different states, which could have constructed a challenge in the classification.

ACKNOWLEDGMENTS

This work is supported by NSF grant #2218077, and the startup fund provided by the Cain Department of Chemical Engineering at Louisiana State University. We also appreciate the Ning Wu lab at Colorado School of Mines, Chemical and Biological Engineering, for providing the experimental datasets for the validation.

AUTHOR DECLARATIONS

Conflict of Interest

The authors declare no competing financial interests.

Author contributions

Andres Lizano-Villalobos: Software conceptualization and development and testing; writing – original draft preparations. **Benjamin Namikas**: software development and testing. **Xun Tang**: Conceptualization; writing – original draft preparations; writing – review and editing; funding acquisition.

DATA AVAILABILITY STATEMENT

Related codes are accessible at Github, via `https://github.com/alizano94/SiNN-StateRep.git`.

[1] A. K. Sarkar, K. Debnath, H. Arora, P. Seth, N. R. Jana, and N. R. Jana, "Direct cellular delivery of exogenous genetic material and protein via colloidal nano-assemblies with biopolymer," ACS Applied Materials Interfaces **14** (January 5, 2022), 10.1021/acsami.1c22009.

[2] I. Coropceanu, E. M. Janke, J. Portner, D. Haubold, T. D. Nguyen, A. Das, C. P. N. Tanner, J. K. Utterback, S. W. Teitelbaum, M. H. Hudson, N. A. Sarma, A. M. Hinkle, C. J. Tassone, A. Eychmüller, D. T. Limmer, M. O. d. l. Cruz, N. S. Ginsberg, and D. V. Talapin, "Self-assembly of nanocrystals into strongly electronically coupled all-inorganic supercrystals," Science **375** (2022-03-25), 10.1126/science.abm6753.

[3] H. Po, C. Dabard, B. Roman, E. Reyssat, J. Bico, B. Baptiste, E. Lhuillier, and S. Ithurria, "Chiral helices formation by self-assembled molecules on semiconductor flexible substrates," ACS Nano **16** (February 2, 2022), 10.1021/acsnano.1c09982.

[4] Longjiang Ding, Xiaoliang Chen, Wenhe Ma, Jiang Li, Xiaoguo Liu, Chunhai Fan, Guangbao Yao, Longjiang Ding, Xiaoliang Chen, Wenhe Ma, Jiang Li, Xiaoguo Liu, Chunhai Fan, and Guangbao Yao, "Dna-mediated regioselective encoding of colloids for programmable self-assembly," Chemical Society Reviews **52** (2023/08/14), 10.1039/D2CS00845A.

[5] L. Guillemeney, L. Lermusiaux, G. Landaburu, B. Wagnon, B. Abécassis, L. Guillemeney, L. Lermusiaux, G. Landaburu, B. Wagnon, and B. Abécassis, "Curvature and self-assembly of semi-conducting nanoplatelets," Communications Chemistry 2022 5:1 **5** (2022-01-12), 10.1038/s42004-021-00621-z.

[6] N. Kang, J. Zhu, X. Zhang, H. Wang, and Z. Zhang, "Reconfiguring self-assembly of photoresponsive hybrid colloids," Journal of the American Chemical Society **144** (March 10, 2022), 10.1021/jacs.2c00432.

[7] B. Ni, G. Gonzalez-Rubio, and H. Cölfen, "Self-assembly of colloidal nanocrystals into 3d binary mesocrystals," Accounts of Chemical Research **55** (June 9, 2022), 10.1021/acs.accounts.2c00074.

[8] A. Pal, C. A. D. Filippo, T. Ito, M. A. Kamal, A. V. Petukhov, C. D. Michele, and P. Schurtenberger, "Shape matters in magnetic-field-assisted assembly of prolate colloids," ACS Nano **16** (February 9, 2022), 10.1021/acsnano.1c09208.

[9] K. Bahri, H. Eslami, and F. Müller-Plathe, "Self-assembly of model triblock janus colloidal particles in two dimensions," Journal of Chemical Theory and Computation **18** (February 14, 2022), 10.1021/acs.jctc.1c01116.

[10] B. Butler and B. Butler, "On the development of tools for the study of colloidal self-assembly," (2024), 10.7302/23058.

[11] K. J. Bishop, S. L. Biswal, B. Bharti, K. J. Bishop, S. L. Biswal, and B. Bharti, "Active colloids as models, materials, and machines," Annual Review of Chemical and Biomolecular Engineering **14** (2023/06/08), 10.1146/annurev-chembioeng-101121-084939.

[12] J. J. Juárez and M. A. Bevan, "Feedback controlled colloidal self-assembly," Advanced Functional Materials **22** (2012/09/25), 10.1002/adfm.201200400.

[13] D. Morphew, J. Shaw, C. Avins, and D. Chakrabarti, "Programming hierarchical self-assembly of patchy particles into colloidal crystals via colloidal molecules," ACS Nano **12** (February 19, 2018), 10.1021/acsnano.7b07633.

[14] X. Tang, B. Rupp, Y. Yang, T. D. Edwards, M. A. Grover, and M. A. Bevan, "Optimal feedback controlled assembly of perfect crystals," ACS Nano **10** (July 11, 2016), 10.1021/acsnano.6b02400.

[15] T. D. Edwards, D. J. Beltran-Villegas, M. A. Bevan, T. D. Edwards, D. J. Beltran-Villegas, and M. A. Bevan, "Size dependent thermodynamics and kinetics in electric field mediated colloidal crystal assembly," Soft Matter **9** (2013/09/11), 10.1039/C3SM50809A.

[16] J. Walker, V. Koutsos, J. Walker, and V. Koutsos, "Spin coating of silica nanocolloids on mica: Self-assembly of two-dimensional colloid crystal structures and thin films," Coatings 2023, Vol. 13, Page 1488 **13** (2023-08-23), 10.3390/coatings13091488.

[17] Y. Xu, "Self-assembled materials based on cellulose nanocrystals and graphene oxide," (2022), 10.14288/1.0406353.

[18] H. Abdi and L. J. Williams, "Principal component analysis," Wiley Interdisciplinary Reviews: Computational Statistics **2** (2010/07/01), 10.1002/wics.101.

[19] M. Greenacre, P. J. F. Groenen, T. Hastie, A. I. D'Enza, A. Markos, E. Tuzhilina, M. Greenacre, P. J. F. Groenen, T. Hastie, A. I. D'Enza, A. Markos, and E. Tuzhilina, "Principal component analysis," Nature Reviews Methods Primers 2022 2:1 **2** (2022-12-22), 10.1038/s43586-022-00184-w.

[20] F. Kherif and A. Latypova, "Principal component analysis," Machine Learning (2020/01/01), 10.1016/B978-0-12-815739-8.00012-2.

[21] Antonia Statt, D. C. Kleeblatt, W. F. Reinhart, Antonia Statt, D. C. Kleeblatt, and W. F. Reinhart, "Unsupervised learning of sequence-specific aggregation behavior for a model copolymer," Soft Matter **17** (2021/08/25), 10.1039/D1SM01012C.

[22] M. Vermeulen, K. Smith, K. Eremin, G. Rayner, and M. Walton, "Application of uniform manifold approximation and projection (umap) in spectral imaging of artworks," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy **252** (2021/05/05), 10.1016/j.saa.2021.119547.

[23] B. Ghojogh, M. Crowley, F. Karray, and A. Ghodsi, "Uniform manifold approximation and projection (umap)," Elements of Dimensionality Reduction and Manifold Learning (2023), 10.1007/978-3-031-10602-6_17.

[24] W. E. Marcilio-Jr, D. M. Eler, F. V. Paulovich, and R. M. Martins, "Humap hierarchihcal uniform manifold approximation and projection," IEEE Transactions on Visualization and Computer Graphics (2024), 10.1109/TVCG.2024.3471181.

[25] Z. Li, Q. Fan, and Y. Yin, "Colloidal self-assembly approaches to smart nanostructured materials," Chemical Reviews **122** (November 8, 2021), 10.1021/acs.chemrev.1c00482.

[26] R. Mao, J. O'Leary, A. Mesbah, and J. Mittal, "A deep learning framework discovers compositional order and self-assembly pathways in binary colloidal mixtures," JACS Au **2** (July 19, 2022), 10.1021/jacsau.2c00111.

[27] "Nonlinear dimensionality reduction," Information Science and Statistics (2007), 10.1007/978-0-387-39351-3.

[28] W. Jia, M. Sun, J. Lian, S. Hou, W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," Complex Intelligent Systems 2022 8:3 **8** (2022-01-21), 10.1007/s40747-021-00637-x.

[29] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, and T. Baker, "Analysis of dimensionality reduction techniques on big data | ieee journals magazine | ieee xplore," IEEE Access **8** (2020), 10.1109/ACCESS.2020.2980942.

[30] M. Dijkstra, E. Luijten, M. Dijkstra, and E. Luijten, "From predictive modelling to machine learning and reverse engineering of colloidal self-assembly," Nature Materials 2021 20:6 **20** (2021-05-27), 10.1038/s41563-021-01014-2.

[31] C. S. Adorf, T. C. Moore, Y. J. U. Melle, and S. C. Glotzer, "Analysis of self-assembly pathways with unsupervised machine learning algorithms," The Journal of Physical Chemistry B **124** (December 8, 2019), 10.1021/acs.jpcb.9b09621.

[32] T. Bereau, D. Andrienko, and K. Kremer, "Research update: Computational materials discovery in soft matter," APL Materials **4** (2016/05/01), 10.1063/1.4943287.

[33] J. Wang and A. L. Ferguson, "Nonlinear machine learning in simulations of soft and biological materials," Molecular Simulation **44** (2018-9-22), 10.1080/08927022.2017.1400164.

[34] J. O'Leary, Runfang Mao, E. J. Pretti, J. A. Paulson, Jeetain Mittal, Ali Mesbah, Jared O'Leary, Runfang Mao, E. J. Pretti, J. A. Paulson, Jeetain Mittal, and Ali Mesbah, "Deep learning for characterizing the self-assembly of three-dimensional colloidal systems," Soft Matter **17** (2021/02/04), 10.1039/D0SM01853H.

[35] R. Luo, W. Liao, X. Huang, Y. Pi, and W. Philips, "Feature extraction of hyperspectral images with semisupervised graph learning | ieee journals magazine | ieee xplore," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **9** (2016), 10.1109/JSTARS.2016.2522564.

[36] J. Wen, Z. Tian, H. She, and W. Yan, "Feature extraction of hyperspectral images based on preserving neighborhood discriminant embedding | ieee conference publication | ieee xplore," 2010 International Conference on Image Analysis and Signal Processing (2010), 10.1109/IASP.2010.5476119.

[37] V. Lotito and T. Zambelli, "Pattern detection in colloidal assembly: A mosaic of analysis techniques," Advances in Colloid and Interface Science **284** (2020/10/01), 10.1016/j.cis.2020.102252.

[38] F. Nina-Paravecino, V. Manian, F. Nina-Paravecino, and V. Manian, "Spherical harmonics as a shape descriptor for hyperspectral image classification," https://doi.org/10.1117/12.850732 **7695** (2010/05/13), 10.1117/12.850732.

[39] V. Varanasi and R. Hegde, "Robust online direction of arrival estimation using low dimensional spherical harmonic features | ieee conference publication | ieee xplore," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2017), 10.1109/ICASSP.2017.7952208.

[40] D. Vranic, D. Saupe, and J. Richter, "Tools for 3d-object retrieval: Karhunen-loeve transform and spherical harmonics | ieee conference publication | ieee xplore," 10.1109/MMSP.2001.962749.

[41] A. Bleakie and D. Djurdjanovic, "Feature extraction, condition monitoring, and fault modeling in semiconductor manufacturing systems," Computers in Industry **64** (2013/04/01), 10.1016/j.compind.2012.10.002.

[42] S. S. Funai and D. Giataganas, "Thermodynamics and feature extraction by machine learning," Physical Review Research **2** (2020-09-15), 10.1103/PhysRevResearch.2.033415.

[43] D. Doan, D. Doan, D. J. Echeveste, D. J. Echeveste, J. Kulikowski, J. Kulikowski, X. W. Gu, and X. W. Gu, "Machine learning analysis of self-assembled colloidal cones," Soft Matter **18** (2022/02/16), 10.1039/D1SM01466H.

[44] M. Schöttle, T. Tran, H. Oberhofer, and M. Retsch, "Machine learning enabled image analysis of time-temperature sensing colloidal arrays," Advanced Science **10** (2023/03/01), 10.1002/advs.202205512.

[45] A. Lizano and X. Tang, "Convolutional neural network-based colloidal self-assembly state classification," Soft Matter **19**, 3450–3457 (2023).

[46] G. Stewart , M. Al-Khassaweneh, G. Stewart , and M. Al-Khassaweneh, "An implementation of the hdbscan* clustering algorithm," Applied Sciences 2022, Vol. 12, Page 2405 **12** (2022-02-25), 10.3390/app12052405.

[47] D. Chicco, "Siamese neural networks: An overview," Methods in Molecular Biology (2021), 10.1007/978-1-0716-0826-5_3.

[48] "The laser-induced damage change detection for optical elements using siamese convolutional neural networks," Applied Soft Computing **87** (2020/02/01), 10.1016/j.asoc.2019.106015.

[49] N. Osiecka-Drewniak, A. Deptuch, M. Urbańska, and E. Juszyńska-Gałązka, "A siamese neural network framework for glass transition recognition," Soft Matter **20**, 2400–2406 (2024).

[50] R. Chiplunkar and B. Huang, "Siamese neural network-based supervised slow feature extraction for soft sensor application | ieee journals magazine | ieee xplore," IEEE Transactions on Industrial Electronics **68** (2021), 10.1109/TIE.2020.3014574.

[51] X. Zhu, Y. Gao, R. Mhana, T. Yang, B. L. Hanson, X. Yang, J. Gong, and N. Wu, "Synthesis and propulsion of magnetic dimers under orthogonally applied electric and magnetic fields," Langmuir **37**, 9151–9161 (2021).

[52] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," Proceedings of the IEEE **109**, 43–76 (2020).

[53] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, "Transfer learning: a friendly introduction," Journal of Big Data **9**, 102 (2022).

[54] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," Computers and Electronics in Agriculture **173**, 105393 (2020).

[55] H. Tang, H. Zhu, L. Fei, T. Wang, Y. Cao, and C. Xie, "Low-illumination image enhancement based on deep learning techniques: a brief review," in *Photonics*, Vol. 10 (MDPI, 2023) p. 198.

[56] X. Shen, Q. Li, Y. Tian, and L. Shen, "An uneven illumination correction algorithm for optical remote sensing images covered with thin clouds," Remote Sensing **7**, 11848–11862 (2015).

[57] J.-L. Starck, M. K. Nguyen, and F. Murtagh, "Wavelets and curvelets for image deconvolution: a combined approach," Signal Processing **83** (2003/10), 10.1016/S0165-1684(03)00150-6.

[58] J. A. Dobrosotskaya and A. L. Bertozzi, "A wavelet-laplace variational technique for image deconvolution and inpainting," IEEE Transactions on Image Processing **17** (2008), 10.1109/TIP.2008.919367.

[59] W.-H. Lee, M. Ozger, U. Challita, and K. W. Sung, "Noise learning-based denoising autoencoder," IEEE Communications Letters **25**, 2983–2987 (2021).