JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

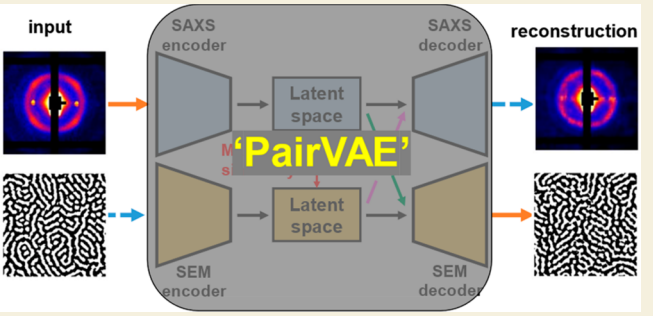
pubs.acs.org/jacsau
Article

Pair-Variational Autoencoders for Linking and Cross-Reconstruction of Characterization Data from Complementary Structural Characterization Techniques

Shizhao Lu and Arthi Jayaraman*

Cite This: JACS Au 2023, 3, 2510−2521
Read Online

ABSTRACT: In materials research, structural characterization often requires multiple complementary techniques to obtain a holistic morphological view of a synthesized material. Depending on the availability and accessibility of the different characterization techniques (e.g., scattering, microscopy, spectroscopy), each research facility or academic research lab may have access to high-throughput capability in one technique but face limitations (sample preparation, resolution, access time) with other technique(s). Furthermore, one type of structural characterization data may be easier to interpret than another (e.g., microscopy images are easier to interpret than small-angle scattering profiles). Thus, it is useful to have machine learning models that can be trained on paired structural characterization data from multiple techniques (easy and difficult to interpret, fast and slow in data collection or sample preparation) so that the model can generate one set of characterization data from the other. In this paper we demonstrate one such machine learning workflow, Pair-Variational Autoencoders (PairVAE), that works with data from small-angle X-ray scattering (SAXS) that present information about bulk morphology and images from scanning electron microscopy (SEM) that present two-dimensional local structural information on the sample. Using paired SAXS and SEM data of newly observed block copolymer assembled morphologies [open access data from Doerk, G. S.; et al. *Sci. Adv.* **2023**, *9* (2), eadd3687], we train our PairVAE. After successful training, we demonstrate that the PairVAE can generate SEM images of the block copolymer morphology when it takes as input that sample's corresponding SAXS 2D pattern and vice versa. This method can be extended to other soft material morphologies as well and serves as a valuable tool for easy interpretation of 2D SAXS patterns as well as an engine for generating ensembles of similar microscopy images to create a database for other downstream calculations of structure−property relationships.



KEYWORDS: *machine learning, material characterization, scattering, microscopy, variational autoencoder*

INTRODUCTION

Machine learning and artificial intelligence are becoming invaluable methods that enable faster discovery, innovation, and automation in chemical and materials sciences and engineering.[1−10] In particular, machine learning has found tremendous use in automating materials' structural analysis, which is a vital step in establishing the design−structure−property relationship of a novel material.[11−21] For example, structural characterization of materials often depends on microscopy imaging techniques (e.g., scanning electron microscopy (SEM), transmission electron microscopy (TEM), or atomic force microscopy (AFM)) to visualize nanoscale or microscale structural features or patterns.[22] Deep learning models used for pattern recognition and image analysis have been adopted as a viable means to automatically extract structural information from microscopy images regarding the ordered arrangements of the molecules, types of ordered assembly, and detection of objects' (e.g., nanoparticles, assembled domains) shapes and sizes.[23−35] There are unique challenges, however, with training machine learning models that are used for everyday photographic image analysis to analyze materials' microscopy images. For example, compared to photographic images of everyday objects that often are easily recognizable, materials' microscopy images require detailed metadata of the material, chemistry, synthesis conditions, and imaging process conditions as labels. The richness of metadata associated with each microscopy image also implies limitation on the curation of large microscopy

Received: May 30, 2023
Revised: July 11, 2023
Accepted: July 11, 2023
Published: July 21, 2023

https://doi.org/10.1021/jacsau.3c00275
JACS Au 2023, 3, 2510−2521

databases because a generic (universal) labeling scheme would be nearly impossible for microscopy images coming from diverse sources. Furthermore, a unique problem to the (nonbiological) materials field is the smaller dataset sizes for the microscopy images collected from samples due to the laborious sample preparation required to acquire each image; this problem restricts the use of deep learning models that require large training datasets like those available in biomedical fields.[36,37] To address this challenge with small datasets of materials' microscopy images , Lu et al.[38] recently proposed a semi-supervised transfer learning workflow that successfully classified the materials' morphologies from TEM images of protein/peptide nanowires despite being trained with fewer than 10 labeled images per morphology.

Another machine learning approach that can be applied to the small and highly chemistry/material/processing-specific microscopy image datasets is *generative* machine learning. Generative machine learning models are trained to reconstruct outputs that resemble the inputs, where the input could be two-dimensional images, three-dimensional structure, graph of a chemical structure of a molecule, $x-y$ plot, etc. One specific type of a generative machine learning model is an autoencoder that can learn reconstruction of complex input data or obtain concise, low-dimensional latent space representations of the complex input data. An autoencoder consists of an encoder that encodes information from the input data to a set of lower-dimensional latent variables and a decoder that decodes information from a set of latent variables back to the same number of dimensions as the input data. An autoencoder is typically trained by minimizing the mismatch (i.e., loss) between the reconstructed data and the input data. Materials scientists have applied autoencoders for multiple purposes such as reconstruction of experimental characterization data,[39−46] molecular structure,[47−50] or microscopy images;[41,46,51,52] clustering[53,54] and/or classification[42,43,45,48] of the latent space representations; obtaining material design parameters[39,40] or deriving order parameters[55−59] from latent space representations; and molecular[47] or material[52] property optimization based on the latent space representations. Some of the studies mentioned here[43−48,50,52,53,57] used a modified version of the autoencoder called a *variational autoencoder* (VAE)[60] that maps the encoded latent space to a multidimensional standard Gaussian distribution, which has the benefit of a continuous latent space compared to the sparse latent space that one would get from the encodings of an unmodified autoencoder. In a recent study, Yaman et al.[46] demonstrated a dual VAE machine learning workflow that can be used to couple the surface photonic response spectroscopy (SPR) profiles and SEM images of gold nanoparticles. They have shown that by correlating the latent space of one VAE trained on SPR with the latent space of another VAE trained on SEM, they can correctly generate (or construct) an SPR profile from a corresponding SEM image or vice versa.

In this article, partly inspired by the work of Yaman et al., we present a machine learning workflow called Pair-Variational Autoencoders (PairVAE) that can enable cross-reconstruction and cross-generation of complementary characterization data of soft materials. As a proof of concept, we demonstrate that PairVAE can reconstruct either small-angle X-ray scattering (SAXS) patterns or SEM images of block copolymer morphologies from the other type of characterization data with a limited number of training data of SAXS−SEM pairs. For interested readers, in Table S1 we compare and contrast our PairVAE work (in terms of purpose, data domain, dataset format, model training protocol) against the work of Yaman et al.[46] that served as an inspiration to us. As our work described in this paper uses block copolymer (BCP) morphologies as an example for demonstrating PairVAE, we discuss briefly BCPs and the types of information that one obtains by conducting SAXS and SEM measurements on them.

BCPs are polymers consisting of two or more types of monomers with repeating sequences of different monomers. The self-assembly or directed assembly of BCPs in films or in solution generates rich ordered morphologies including spheres, cylinders, lamellae, gyroids, vesicles, and other bicontinuous structures.[61−64] For melts of a simple linear diblock copolymer with two immiscible monomer chemistries, past studies using theory, simulations, and experiments have all shown self-assembly into four canonical ordered BCP morphologies—lamella, cylinder, spheres, and gyroid—by varying the BCP composition and/or temperature or the Flory−Huggins χ parameter (an indicator of immiscibility of the two monomers in the BCP).[65−67] In experiments, these BCP morphologies are most commonly characterized by complementary experimental techniques: SAXS or small-angle neutron scattering (SANS) and TEM or SEM. The full two-dimensional SAXS pattern quantifies the degree of positional and orientational order present in the BCP morphology over a range of length scales (inverse wavevector q). The azimuthally averaged one-dimensional SAXS or SANS scattering intensity $I(q)$ versus scattering wavevector q is used to quantify the BCP domain spacing. In contrast, SEM or TEM images provide a 2D visualization of BCP morphologies at specific length scales. Unlike the SAXS and SANS patterns (2D or 1D), which provide average structural information on the bulk sample, SEM or TEM images for the same material can differ from sample to sample due to selection of the imaging region, differences of magnification/contrast, and existence of defects in the domains over the sample. While SAXS has the benefit of outputting statistical average or bulk structure information over multiple length scales, the data in wavevector q space are abstract and require nontrivial work in interpretation using analytical models[68,69] or advanced computational techniques.[15,19,70] Unlike SAXS or SANS profiles, which are harder to analyze and prone to incorrect interpretation if one fits with analytical models that are either approximate or inapplicable for the structure at hand, the SEM or TEM images are a real-space visualization and less prone to incorrect interpretation. Thus, having both sets of results can provide an unambiguous interpretation of the structure in the BCP system at various length scales. One could apply fast Fourier transform on an SEM image to obtain a SAXS-like scattering pattern, albeit with worse statistics than that obtained from SAXS measurement; furthermore, generation of local real-space SEM morphology representations from a SAXS pattern has not been reported in the literature. Therefore, having a machine learning model that can couple/pair complementary SAXS and SEM results will be valuable, as it can overcome the individual limitations/challenges for either approach. We describe our machine learning workflow, PairVAE, for linking and cross-reconstruction of 2D SAXS patterns and SEM images that provide complementary information about soft materials' morphologies.

This article is organized as follows. In Method, we describe the development of PairVAE, including the description of the data used to train the models. In the Results and Discussion,

https://doi.org/10.1021/jacsau.3c00275
JACS Au 2023, 3, 2510−2521

we describe the performance of the trained models and reasons for the observed trends. In the Conclusion, we provide a summary of the outcomes and a list of potential applications of the PairVAE method broadly, beyond the case described in this article.

METHOD

Our machine learning workflow PairVAE for cross-reconstruction of two types of complementary characterizations, 2D SAXS patterns and SEM images, was trained on paired SAXS and SEM data of BCP morphologies reported recently by Doerk et al.[71] They used blends of two BCPs, both with polystyrene (PS) and poly(methyl methacrylate) (PMMA) blocks but with varying composition; one BCP was a lamella-forming symmetric PS-*b*-PMMA, and the other BCP was a cylinder-forming PMMA-minority PS-*b*-PMMA.[72] These BCP blends were assembled into unique morphologies using template-directed assembly by varying the width of the substrate pattern.[71] For each of these morphologies, the authors presented complementary characterization data in the form of 2D SAXS patterns and SEM images. Figure 1 shows a representative 2D SAXS pattern and SEM image for each of the four morphologies they observed. BCP structures seen in SEM images were partially infiltrated with aluminum oxide to enhance visual contrast and may have imparted imperfections and defects in the processing. The "hexagonal" morphology consists of cylinder-like domains (bright dots and dim vertical lines) and lamella-like domains (bright vertical lines) in the SEM images with regular domain spacing. The authors noted that the "skew", "alternating lines", and "ladders" morphologies were not seen in previous studies. The "skew" morphology consists of bilayers of lamella-like domains at the air interface with cylinder-like domains at the template interface. The "alternating lines" morphology consists of vertical intercalating lines of bright lamella-like domains and dim cylinder-like domains, while the "ladders" morphology consists of oblong shapes with lamella-like domains parallel to the template grating direction and cylinder-like domains orthogonal to the template grating direction. In their work, SAXS was the high-throughput technique while the SEM characterization was slower and laborious and considered a bottleneck in their production of this paired characterization data. Their open-access dataset is tremendously useful for training and testing of a machine learning workflow like our PairVAE that can cross-reconstruct SEM (bottleneck method but easier to interpret) images from SAXS (high-throughput method but harder to interpret) data and vice versa.

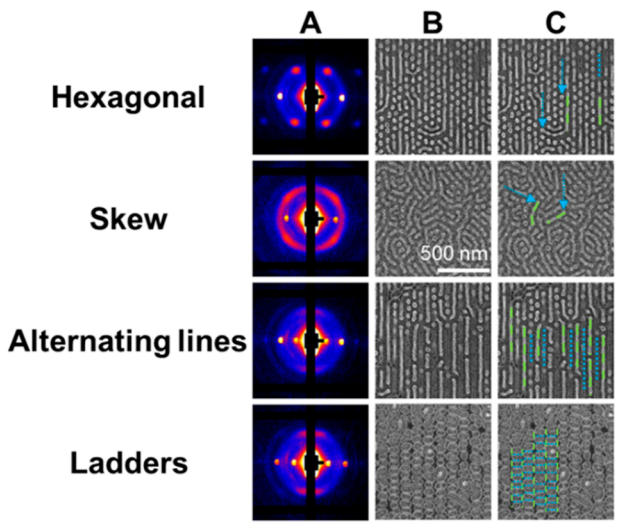


Figure 1. Representative characterization data of four different morphologies observed with a mixture of lamella- and cylinder-forming PS-*b*-PMMA BCPs assembled on a chemical grating template. The images in columns **A**, **B**, and **C** are example SAXS patterns, randomly cropped SEM images, and annotated SEM images for each of the four morphologies. In the annotated images, green dashed lines indicate examples of lamella-like domains, and blue arrows and dotted lines indicate cylinder-like domains. The white scale bar denoting 500 nm (in "skew" panel B) applies to all SEM images without scale bars. The full SEM images of the four morphologies are presented in Figure S1. All SAXS and full SEM images were obtained from the open-access block copolymers characterization dataset by Doerk et al.[71] from the Materials Data Facility.[73,74] The SAXS dataset was obtained from DOI: 10.18126/qboh-6fav and the SEM dataset from DOI: 10.18126/09ow-8g8k. This image is adapted from ref71. CC BY 4.0.

PairVAE is composed of two sets of variational autoencoders (VAEs). Before the two sets of encoders and decoders in the two VAEs are trained together to handle cross-reconstruction, each VAE is first trained on one type of characterization data through solo training, as shown in Figure 2A. We first briefly describe the dataset and the preprocessing of the characterization data used in these training steps.

We obtained from the open-access data depository 72 pairs of SAXS and SEM characterization data that were imaged at the same position in the larger grating template with the knowledge that the perception field of SAXS is larger than that of SEM. For SAXS data, we use the 2D SAXS pattern images resized to 192 pixels × 192 pixels for training and reconstruction. For SEM data, the full SEM images are 830 pixels × 1280 pixels, with examples shown in Figure S1. However, we do not use the full SEM images as our input for two reasons. One reason is that the different morphologies are visually distinguishable at smaller regions than those in the full SEM image but less distinguishable with the full SEM image. The second reason is that the convolutional neural network (CNN) models that we use to process images work well with image sizes of approximately 256 pixels × 256 pixels and smaller. We preprocess the full SEM images with adaptive thresholding[75] to obtain binarized versions of the SEM images with better delineation of the PMMA domains versus PS domains. We have found adaptive thresholding to be much better at delineation of PMMA versus PS domains of SEM images exhibiting the "alternating lines" or "ladders" morphology compared to other thresholding methods described in Figure S2.

During training of the SEM encoder and decoder, we randomly crop a 192 × 192 section of the full SEM image at each training epoch and reconstruct the cropped section. Random cropping is an effective data augmentation approach for our small-size dataset. For SEM images, we do not expect pixel-to-pixel exact reconstruction from our decoder based on our knowledge of the high statistical variance of image representations belonging to the same morphology. For SEM images, we instead aim for morphologically similar reconstructions of the input image and apply quantitative metrics to quantify the similarity in morphology: the two-point correlation function[76] and the lineal path function.[77] Our choice of these two morphological quantification metrics was

https://doi.org/10.1021/jacsau.3c00275
JACS Au 2023, 3, 2510−2521

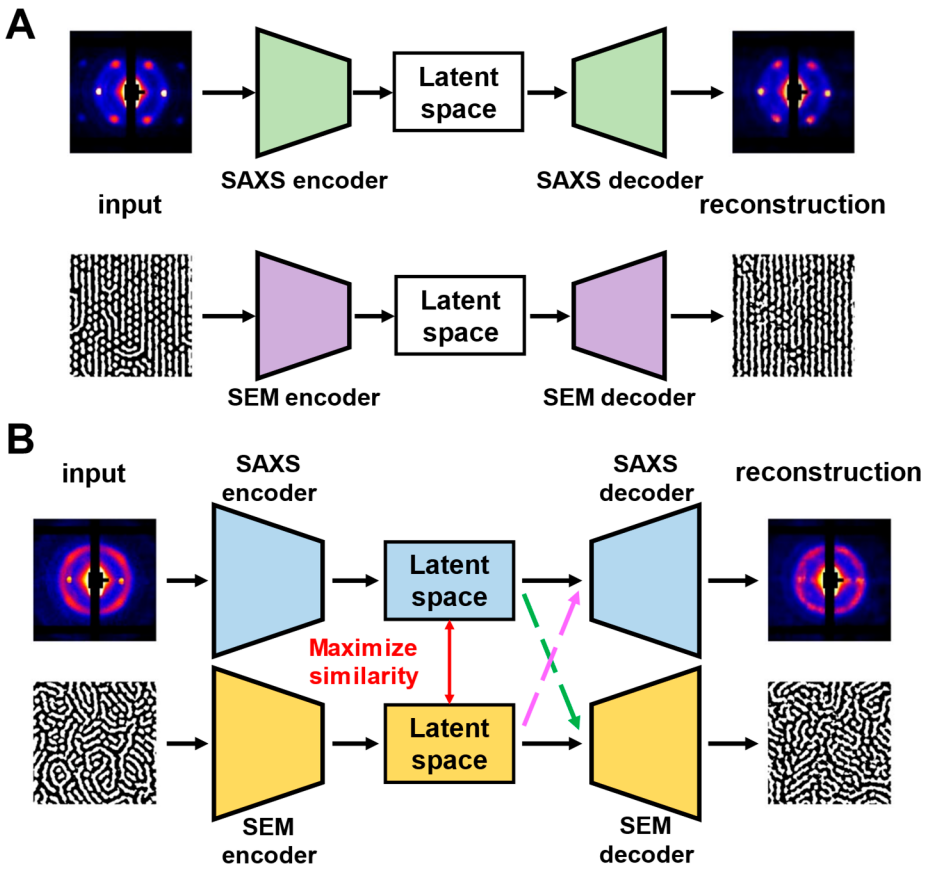


Figure 2. Schematics of the PairVAE workflow. (A) First, solo training of VAEs for SAXS and SEM images separately is performed. (B) Pair training of VAEs with SAXS and SEM images concurrently with additional loss functions beyond those used in the solo VAEs is performed. PairVAE has four reconstruction capabilities: SAXS profiles from SAXS input (denoted as SAXS → SAXS in the text), SAXS profiles from SEM input (i.e., SEM → SAXS), SEM images from SEM input (SEM → SEM), and SEM images from SAXS input (SAXS → SEM).

inspired by recent works on rock image reconstruction.[78,79] One can also use other morphology metrics such as the radial distribution function, orientational order parameter, graph connectivity, etc. for morphological quantification. The two-point correlation function $S(r)$ gives us both the pixel density of the PMMA domains at various positions (radial distance r) and the degree of positional order through the oscillations in $S(r)$. The linear path function gives us the probability distribution of the length of line segments in PMMA domains along a path. If done along the vertical axis, the lineal path function indicates the degree of connectivity of the PMMA domains in the vertical direction; similarly, a lineal path function can be done along the horizontal path. These two metrics are only used as a performance assessment of the reconstructed SEMs (after a model has been trained) and were not calculated or inferred in any way during training. The morphological metrics were calculated with the PoreSpy package.[80] A lower-dimensional latent space is obtained through solo training of each VAE as well as reconstructions of the input characterization data.

In the second part, i.e., pair training, we train the encoders and decoders of the two VAEs in tandem by maximizing the similarity of the two latent spaces, as shown in Figure 2B. Additional cross-reconstruction training losses are established to enable cross-reconstruction capabilities of one type of characterization data from another. We note that while solo training involves reconstruction of SAXS or SEM in their respective space, pair training attempts to bridge the two different data spaces and tackle the many-to-one (SEM-to-SAXS) and one-to-many (SAXS-to-SEM) cross-reconstruction problems. In selection of training protocols, we found that different from solo training, random cropping of SEM images every epoch during training renders the SEM decoder untrainable, as shown by the unabating losses in Figure S3D,E. Therefore, we adopt a two-part training protocol for pair training, with a first part of multiple random cropping SEM sessions and a second part focusing on reconstruction of one randomly cropped SEM instance. A detailed description of the VAE models employed and training protocols of both solo VAE training and pair VAE training is given in the Supporting Information.

RESULTS AND DISCUSSION

First, in Figure 3A we show the results of solo-trained VAEs (i.e., a VAE generating an SEM image from an SEM image and a VAE generating a SAXS pattern from a SAXS pattern). We include one example of the true and reconstructed SAXS pattern and SEM image for each of the four morphologies obtained with solo-trained VAEs; we note that the selected images are from validation sets and were not seen during training. We observe that the SAXS reconstructions using the solo-trained SAXS VAE exhibit a good match with their true counterparts. For SEM reconstructions, we observe only partial resemblance to the true SEM images. For the "skew" and "ladders" morphologies, their SEM reconstructions are visually different than their true SEM images. In Figure 3B, we show

https://doi.org/10.1021/jacsau.3c00275
JACS Au 2023, 3, 2510−2521

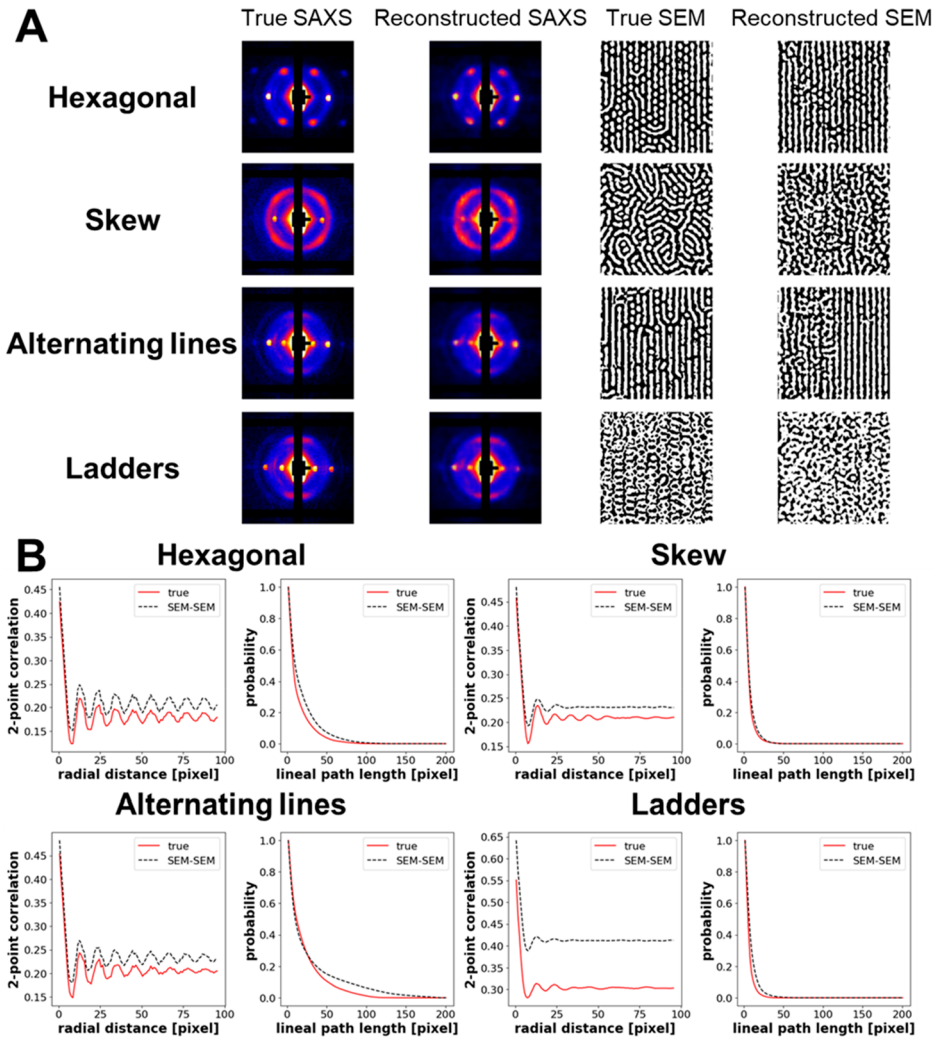


Figure 3. Reconstruction performance of solo-trained VAEs. (A) Examples of SAXS and SEM reconstructions for images from the validation set with solo-trained VAEs. (B) Variation of two-point correlation functions $S(r)$ of PMMA domains along radial distance and vertical lineal path functions of PMMA domains.

the calculated $S(r)$ of PMMA domains along the radial distance and the vertical lineal path function of PMMA domains for each true and reconstructed SEM image in Figure 3A. In our case, as the chemical grating direction is vertical, the vertical lineal path function has more features. The horizontal lineal path functions are provided in Figure S4. For all four morphology examples, the reconstructed SEM pixel density of PMMA domains is higher by a small margin, with that of the "ladders" morphology being the most different. Reconstructions of both "hexagonal" and "alternating lines" morphologies can capture the spatial order of the PMMA domains, indicated as having a similar period of oscillation in $S(r)$ as the $S(r)$ of the true SEM images, but give slightly higher lineal path function compared to that of the true SEM images. For the "skew" morphology, the reconstruction does not capture the longer-range positional order of the true SEM image. As the "ladders" morphology has more statistical variance in the sample, it is not surprising that it is harder to reconstruct the "ladders" morphology than the other three morphologies. Additionally, unlike the other three morphologies, where the PMMA domains occupy full line segments or dots, in the "ladders" morphology the PMMA domains reside on the periphery of the ladders. Furthermore, the highly nonuniform shape of the ladders indicates that the domains formed by different ladder peripheries are also nonuniform.

Next, in Figure 4A, we show the reconstructions from the pair-trained VAE of the same examples that we showed in Figure 3A (which were obtained with solo-trained VAEs). We observe in Figure 4A that SAXS reconstruction from SAXS input (denoted as "SAXS → SAXS") with the pair-trained VAE is worse than that using the solo-trained SAXS VAE (Figure 3A). We also observe that SAXS reconstruction from SEM input (denoted as "SEM → SAXS") is comparable to SAXS reconstruction from SAXS input (i.e., SAXS → SAXS) with the pair-trained VAE. For SEM reconstruction from SAXS input (denoted as "SAXS → SEM"), for all four morphologies, the reconstructed SEM images are visually more similar to the true SEM images than the SEM → SEM reconstructions with the pair-trained VAE. In general, Figure 4A shows that by pairing SAXS with SEM in pair-trained VAE, we have improved the SEM reconstruction and deteriorated the SAXS reconstruction compared to their solo-trained VAEs (Figure 3A). This unexpected result will be explained below based on latent space complexity in the paragraph after the discussion of the results in Figure 4B.

https://doi.org/10.1021/jacsau.3c00275
JACS Au 2023, 3, 2510−2521

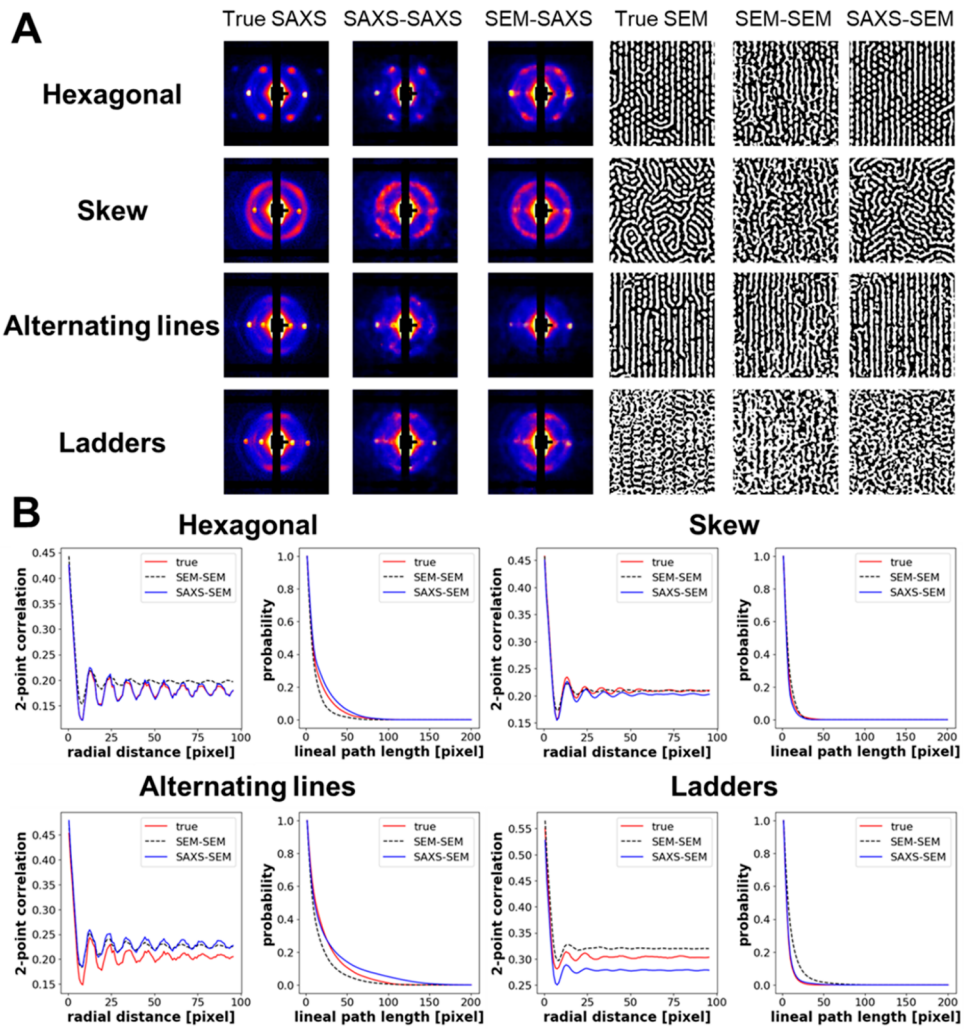


Figure 4. Reconstruction performance of a pair-trained VAE. (A) Examples of SAXS and SEM reconstructions for images from the validation set with a pair-trained VAE. (B) Variation of two-point correlation functions of PMMA domains along radial distance and vertical lineal path functions of PMMA domains.

While the conclusions from Figure 4A are purely based on visual comparisons, in Figure 4B we provide a quantitative comparison of the SEM reconstructions against the true SEM image for both SEM → SEM and SAXS → SEM using the two metrics described in the Method: the two-point correlation function and the vertical linear path function. The horizontal lineal path functions are provided in Figure S5.

For the "hexagonal" morphology, the two-point correlation function $S(r)$ for SAXS → SEM reconstruction closely matches the $S(r)$ of the true SEM image, while the SEM → SEM reconstruction has slightly higher $S(r)$ and does not capture the shape of the $S(r)$ of the true SEM image. Similarly, the vertical lineal path function of the SAXS → SEM reconstruction is closer to the true SEM image than the SEM → SEM is. However, the vertical lineal path function of SAXS → SEM reconstruction is slightly higher than the true SEM image due to more occurrences of lines than in the true SEM image. The SEM → SEM reconstruction's vertical lineal path function is lower than that of the true SEM image due to more broken-up line segments in the reconstructed image.

For the "skew" morphology, the $S(r)$ of the SAXS → SEM reconstruction follows the oscillations of $S(r)$ of the true SEM image closely, albeit with slightly lower values indicating slightly lower pixel density of PMMA domains in the SAXS → SEM reconstruction. The $S(r)$ of the SEM → SEM reconstruction captures the mean of the oscillating $S(r)$ well but does not capture the shape of the $S(r)$. Both SEM → SEM and SAXS → SEM reconstructed "skew" images have a lineal path function closely matching that of the true SEM image.

For the "alternating lines" morphology, the $S(r)$ of both SEM → SEM and SAXS → SEM reconstructions have an oscillating period like that of the true SEM image. However, both reconstructed $S(r)$ curves are higher than those of the true SEM image. Commenting on the reconstructions of the "alternating lines" morphology in Figure 4A, the mismatch in the $S(r)$ is likely due to the inability to accurately reconstruct the minor dotted region in the true SEM image. The lineal path function of the SAXS → SEM reconstruction shows a larger proportion of longer line segments while that of the SEM → SEM reconstruction gives smaller probability values overall due to more broken-up line segments.

For the "ladders" morphology, the $S(r)$ of both SEM → SEM and SAXS → SEM reconstructions are still off from that of the true SEM image but much better than that of the solo-trained SEM reconstruction shown in Figure 3B. The lineal

https://doi.org/10.1021/jacsau.3c00275
JACS Au 2023, 3, 2510−2521

path function of the SAXS → SEM reconstruction is a close match with that of the true SEM image.

The takeaway message so far is that pair training enables both SAXS → SEM and SEM → SAXS cross-reconstruction. Compared with solo-trained VAEs (results in Figure 3), the pair-trained VAE (results in Figure 4) improves SEM reconstruction performance while the SAXS reconstruction performance becomes slightly worse. The trade-off in the reconstruction performance for the two types of characterization data may be explained by the relative complexity of the two latent spaces. The SAXS data that we use have limited modes or representations, making its latent space simpler or more clustered than that of SEM data due to the large statistical variance of PMMA domain shapes and positions in SEM images belonging to the same morphology. Through similarity matching in pair training of the SAXS data of simpler or more clustered latent space with SEM data of sparser latent space, the SAXS latent space becomes sparser (leading to worse reconstruction) and the SEM latent space becomes more convergent (leading to better reconstruction).

After the above detailed analysis of selected examples of solo-trained VAE and pair-trained VAE reconstructions for each morphology, we show the statistical averages and standard deviations of the two-point correlation function for all SEM true images and reconstructions for solo-trained VAE (Figure 5A) and pair-trained VAE (Figure 5B). In Figure 5A,

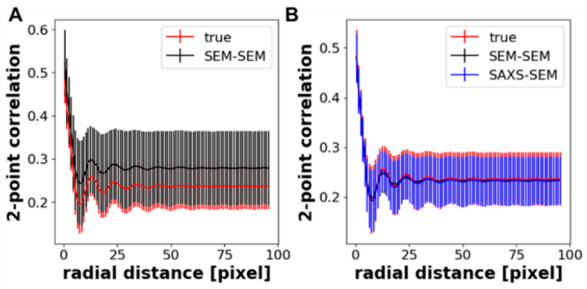


Figure 5. Statistical averages and standard deviations of two-point correlation functions of PMMA domains of all true SEM images (training and validation set) and reconstructions from (A) the solo-trained VAE and (B) the pair-trained VAE.

the solo-trained VAEs' SEM reconstructions give higher two-point correlation function values on average than the true SEM images. In Figure 5B, the statistical average and variance of two-point correlation function of both pair-trained VAEs' SEM reconstructions are a close match with that of the true SEM images. We provide lineal path functions calculated for both solo- and pair-trained VAE reconstructions in Figure S6. We conclude that pair-trained VAE improves SEM reconstructions on average for all SEM images compared to a solo-trained VAE.

We also perform the same solo and pair training on the SAXS−SEM paired dataset but with a different number of training set images and with five different selections of the training set and validation set images (determined by a different random seed) at the same training set size to assess the ability to generalize and illustrate the data efficiency of our workflow. We use the mean absolute error (MAE) between the pixel intensity values of a true SAXS image and of a SAXS reconstruction as the performance quantification metric for the SAXS reconstruction, and we use the MAE between the $S(r)$ of a true SEM image and of a SEM reconstruction as the performance quantification metric for SEM reconstruction. We also use ten random cropped instances from each full SEM image as true SEM images per SAXS image to account for the higher statistical variance of SEM reconstructions as well as for SEM−SAXS reconstructions. The box plots of distributions of all MAEs for SAXS profile reconstructions are shown in Figure 6A. We observe that the median MAE of solo-trained SAXS → SAXS reconstruction is much lower than those of the pair-trained SAXS → SAXS and SEM → SAXS reconstructions; this is consistent with what we observed qualitatively from Figure 3 to Figure 4 that SAXS reconstruction is the best with the solo-trained VAE. The median MAE of pair-trained SEM → SAXS is always a little higher than that of pair-trained SAXS → SAXS. We attribute the observed difference to the statistical variability of the SEM images reflecting local morphology that may not be the mode of morphology that the SAXS captures. We also note that the MAEs of all three reconstructions plateau from 48 training set images to 64 training set images, implying that we do not improve the reconstruction efficacy that much by using training sets larger than 48.

The box plots of distributions of all MAEs for SEM image reconstruction are shown in Figure 6B. We observe that the median MAE of pair-trained SAXS → SEM reconstruction is always lower than those of the other two reconstructions. The performance difference can be explained by our hypothesis that the more clustered latent space of the SAXS images helps converge the SEM latent space to improve SEM reconstruction. The solo- and pair-trained SEM → SEM reconstructions show comparable MAEs for the different training set sizes. SEM reconstructions show improvement with more training set images.

We further probe the capability of PairVAE to generate an ensemble of morphologically similar SEM images directly from the latent space through a trained decoder. We note that we have chosen VAE instead of another deep learning model architecture called U-Net,[81] which is the go-to model for biomedical image segmentation. U-Net has proven to be good for segmentation, a pixelwise reconstruction, due to the multiple skip connections established between the encoder and decoder at multiple hierarchical levels. However, because of the skip connections, the encoder and decoder are trained as a couple in U-Net and cannot be decoupled. Being able to separate the trained encoder and decoder in the VAE model enables us to conduct pair training and explore not just reconstruction but also the image generation ability of the decoder.

To demonstrate the image generation ability of the decoder, we sample latent variables from a multivariate Gaussian distribution constructed with the values of the SAXS latent variables of each morphology example (obtained with the pair-trained SAXS encoder) shown in Figure 4A as the average and the variance of the latent variables of the full SAXS dataset as the variance. We then use the pair-trained SEM decoder to generate images from the sampled latent variables; five generated examples are shown in Figure 7A for each morphology. The generated images present degrees of statistical variations of junctions, vacancies, and defects of PMMA domains that may occur in experimental BCP assembly. When compared to the SAXS → SEM reconstructed images, these generated images (sample size is 100 for each morphology) are morphologically similar as verified visually and by the calculated two-point correlation functions shown in Figure 7B. By changing the magnitude of the variance of the

https://doi.org/10.1021/jacsau.3c00275
JACS Au 2023, 3, 2510−2521

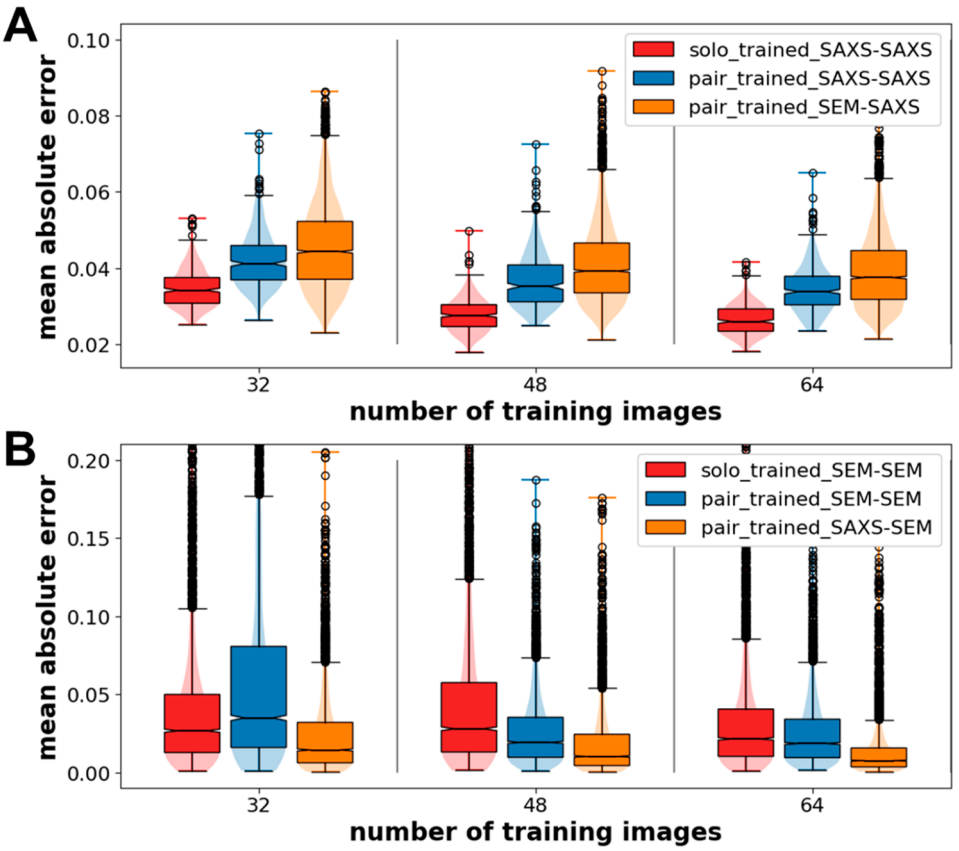


Figure 6. Solo- and pair-trained VAE performance assessment using mean absolute error as the metric (varying the number of training images). (A) Mean absolute error between each pairing of true and reconstructed SAXS images for all images and multiple VAEs trained with five random seeds. The sample size of the box plot is 360 for solo-trained VAE reconstructed SAXS → SAXS and pair-trained VAE reconstructed SAXS → SAXS and 3600 for pair-trained VAE reconstructed SEM → SAXS. (B) Mean absolute error between each pairing of the two-point correlation functions of PMMA domains of true and reconstructed SEM images for all images and multiple VAEs trained with five random seeds (we applied 10 random crops to each full SEM image during performance assessment). The sample size of the box plots is 3600. The median values of both plots are presented in Table S2. The notch of the box plot indicates the 95% confidence interval around the median.

constructed multivariate Gaussian distribution, we observe either better resemblance (smaller variance, Figure S7) or less resemblance (larger variance, Figure S8) to the SAXS → SEM reconstructions.

CONCLUSION

We have developed a novel machine learning workflow called PairVAE that enables cross-reconstruction of complementary material structural characterization data. Here we demonstrated a proof-of-concept application of PairVAE using complementary characterization data, 2D SAXS patterns and SEM images, of ordered block copolymer morphology. By pairing up the latent spaces of SAXS and SEM, PairVAE enables cross-reconstruction of SAXS profiles from the SEM input or SEM images from the SAXS input. We generated morphologically similar SEM images from SAXS input with close match of the two-point correlation functions of the PMMA domains to that of the true SEM images for all data on average. We provided an explanation for the performance trade-off of pair-trained VAE seen for SAXS reconstructions (worse than solo-trained VAE) and SEM reconstructions (better than solo-trained VAE). The explanation is that by pairing the SEM latent space (relatively sparse) with the SAXS latent space (relatively clustered), the SEM latent space becomes more convergent, yielding morphologically closer reconstructions than solo training, while the SAXS latent space becomes less clustered, yielding more unphysical SAXS patterns than solo training. Our PairVAE implementation for SAXS−SEM reconstruction specifically incorporated random cropping of SEM images during training as a means of data augmentation that helps mitigate the small-data handicap. Lastly, we have also shown that generation of ensembles of morphologically similar SEM images is possible by sampling latent variables directly; this is possible only because we can decouple the trained encoder and decoder, which is not possible with more commonly used image segmentation models like U-Net.[81]

The impact of this PairVAE workflow is in the many potential applications it offers, some of which we list below:

(1) Generation of bottleneck characterization data (e.g., laborious sample preparation for microscopy) from high-throughput data (e.g., scattering profile collection) that have complementary purposes.

(2) Generation of easier-to-interpret characterization data (e.g., microscopy image) from harder-to-interpret characterization data (e.g., scattering data) in a data-driven manner without intermediate manual interpretation.

(3) Generation of ensembles of structure or morphology data with designed morphological features for downstream calculation or simulation of other physical

https://doi.org/10.1021/jacsau.3c00275
JACS Au 2023, 3, 2510−2521

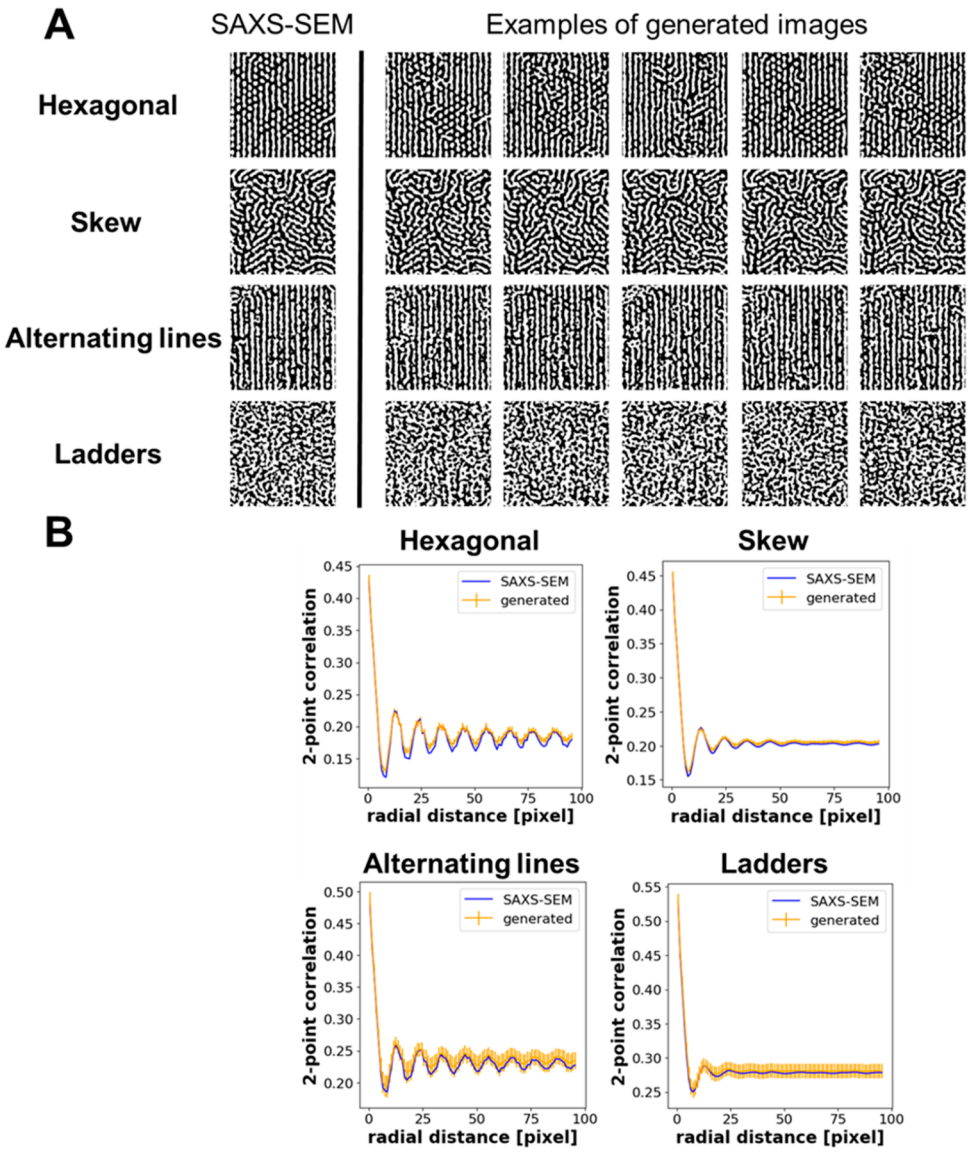


Figure 7. Performance assessment of the generative ability of a pair-trained VAE's SEM decoder. (A) SAXS → SEM reconstruction and five examples of generated SEM images. (B) Statistical averages and standard deviations of two-point correlation functions of PMMA domains of 100 generated SEM images for each morphology example plotted against that of the SAXS → SEM reconstruction.

properties such as optical, electrical, or rheological properties.

(4) Integration with time series data for prediction of time evolution of structure or morphology or other characterization data or properties.

(5) For scattering and microscopy specifically, direct integration with microscopy and scattering experimental setups for real-time high-throughput multipurpose characterization of material structure or morphology or in situ collection and generation of movies of structural characterizations.

■ ASSOCIATED CONTENT

Data Availability Statement

The Google Colab notebooks containing the implementation of the PairVAE workflow are available at https://github.com/arthijayaraman-lab/PairVAE.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.3c00275.

Description of the block copolymer SAXS−SEM morphology characterization dataset, image data preprocessing procedures, python packages utilized and the usages of each package, the variational autoencoder model, the solo training protocol, the pair training protocol, and additional lineal path function figures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Arthi Jayaraman − *Department of Chemical and Biomolecular Engineering and Department of Materials Science and Engineering, University of Delaware, Newark, Delaware 19716, United States;* orcid.org/0000-0002-5295-4581; Email: arthij@udel.edu

https://doi.org/10.1021/jacsau.3c00275
JACS Au 2023, 3, 2510−2521

Author

Shizhao Lu − *Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware 19716, United States;* orcid.org/0000-0003-3055-1877

Complete contact information is available at: https://pubs.acs.org/10.1021/jacsau.3c00275

Author Contributions

CRediT: **Shizhao Lu** conceptualization, data curation, formal analysis, methodology, validation, visualization, writing-original draft; **Arthi Jayaraman** funding acquisition, project administration, resources, supervision, visualization, writing-original draft, writing-review & editing.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge financial support from the U.S. Department of Energy, Office of Science (Grant DE-SC 0023264) for this work and the collaborative discussions with Quentin Michaudel and Ryan Hayward. We are also grateful for the useful technical discussions with Todd Emrick and Jessica Schiffman at the University of Massachusetts, Amherst, that were facilitated via an ongoing collaboration supported by the NSF (DMREF Grant 1921871). We are truly grateful to Gregory S. Doerk, Kevin Yager, and their coauthors for making their block copolymer characterization data openly available, which enabled S.L. to train the models used in PairVAE. The authors also thank Gregory S. Doerk, Jonathan A. Malen, Austin M. Evans, Brent S. Sumerlin, Pramod Reddy, Vasudevan Venkateshwaran, and Rama K. Vasudevan for their questions, comments, and discussion when we presented this method to them for their critique and feedback.

■ REFERENCES

(1) Agrawal, A.; Choudhary, A. Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. *APL Mater.* **2016**, *4* (5), 053208.

(2) Angermueller, C.; Pärnamaa, T.; Parts, L.; Stegle, O. Deep learning for computational biology. *Mol. Syst. Biol.* **2016**, *12* (7), 878.

(3) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361* (6400), 360−365.

(4) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559* (7715), 547−555.

(5) Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W. Deep learning for molecular design—a review of the state of the art. *Mol. Syst. Des. Eng.* **2019**, *4* (4), 828−849.

(6) Kadulkar, S.; Sherman, Z. M.; Ganesan, V.; Truskett, T. M. Machine Learning-Assisted Design of Material Properties. *Annu. Rev. Chem. Biomol. Eng.* **2022**, *13*, 235−254.

(7) Ferguson, A. L.; Brown, K. A. Data-driven design and autonomous experimentation in soft and biological materials engineering. *Annu. Rev. Chem. Biomol. Eng.* **2022**, *13*, 25−44.

(8) Choudhary, K.; DeCost, B.; Chen, C.; Jain, A.; Tavazza, F.; Cohn, R.; Park, C. W.; Choudhary, A.; Agrawal, A.; Billinge, S. J.; et al. Recent advances and applications of deep learning methods in materials science. *npj Comput. Mater.* **2022**, *8* (1), 59.

(9) Patel, R. A.; Webb, M. A. Data-Driven Design of Polymer-Based Biomaterials: High-throughput Simulation, Experimentation, and Machine Learning. *ACS Appl. Bio Mater.* **2023**, DOI: 10.1021/acsabm.2c00962.

(10) Martin, T. B.; Audus, D. J. Emerging Trends in Machine Learning: A Polymer Perspective. *ACS Polym. Au* **2023**, *3* (3), 239−258.

(11) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure-Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* **2012**, *112* (5), 2889−2919.

(12) Ma, R.; Liu, Z.; Zhang, Q.; Liu, Z.; Luo, T. Evaluating Polymer Representations via Quantifying Structure-Property Relationships. *J. Chem. Inf. Model.* **2019**, *59* (7), 3110−3119.

(13) Yuan, S.; Jiao, Z.; Quddus, N.; Kwon, J. S., II; Mashuga, C. V. Developing Quantitative Structure-Property Relationship Models To Predict the Upper Flammability Limit Using Machine Learning. *Ind. Eng. Chem. Res.* **2019**, *58* (8), 3531−3537.

(14) Barnard, A.; Opletal, G. Predicting structure/property relationships in multi-dimensional nanoparticle data using t-distributed stochastic neighbour embedding and machine learning. *Nanoscale* **2019**, *11* (48), 23165−23172.

(15) Beltran-Villegas, D. J.; Wessels, M. G.; Lee, J. Y.; Song, Y.; Wooley, K. L.; Pochan, D. J.; Jayaraman, A. Computational reverse-engineering analysis for scattering experiments on amphiphilic block polymer solutions. *J. Am. Chem. Soc.* **2019**, *141* (37), 14916−14930.

(16) Amamoto, Y. Data-driven approaches for structure-property relationships in polymer science for prediction and understanding. *Polym. J.* **2022**, *54* (8), 957−967.

(17) Liu, Y.; Kelley, K. P.; Vasudevan, R. K.; Funakubo, H.; Ziatdinov, M. A.; Kalinin, S. V. Experimental discovery of structure-property relationships in ferroelectric materials via active learning. *Nat. Mach. Intell.* **2022**, *4* (4), 341−350.

(18) Tao, L.; Byrnes, J.; Varshney, V.; Li, Y. Machine learning strategies for the structure-property relationship of copolymers. *iScience* **2022**, *25* (7), 104585.

(19) Heil, C. M.; Patil, A.; Dhinojwala, A.; Jayaraman, A. Computational Reverse-Engineering Analysis for Scattering Experiments (CREASE) with Machine Learning Enhancement to Determine Structure of Nanoparticle Mixtures and Solutions. *ACS Cent. Sci.* **2022**, *8* (7), 996−1007.

(20) Wu, Z.; Jayaraman, A. Machine Learning-Enhanced Computational Reverse-Engineering Analysis for Scattering Experiments (CREASE) for Analyzing Fibrillar Structures in Polymer Solutions. *Macromolecules* **2022**, *55* (24), 11076−11091.

(21) Heil, C. M.; Ma, Y.; Bharti, B.; Jayaraman, A. Computational Reverse-Engineering Analysis for Scattering Experiments for Form Factor and Structure Factor Determination ("$P(q)$ and $S(q)$ CREASE"). *JACS Au* **2023**, *3* (3), 889−904.

(22) Goodhew, P. J.; Humphreys, J.; Beanland, R. *Electron Microscopy and Analysis*; CRC Press, 2000.

(23) Bostanabad, R.; Zhang, Y.; Li, X.; Kearney, T.; Brinson, L. C.; Apley, D. W.; Liu, W. K.; Chen, W. Computational microstructure characterization and reconstruction: Review of the state-of-the-art techniques. *Prog. Mater. Sci.* **2018**, *95*, 1−41.

(24) Ge, M.; Su, F.; Zhao, Z.; Su, D. Deep learning analysis on microscopic imaging in materials science. *Mater. Today Nano* **2020**, *11*, 100087.

(25) Spurgeon, S. R.; Ophus, C.; Jones, L.; Petford-Long, A.; Kalinin, S. V.; Olszta, M. J.; Dunin-Borkowski, R. E.; Salmon, N.; Hattar, K.; Yang, W.-C. D.; et al. Towards data-driven next-generation transmission electron microscopy. *Nat. Mater.* **2021**, *20* (3), 274−279.

(26) Baskaran, A.; Kautz, E. J.; Chowdhary, A.; Ma, W.; Yener, B.; Lewis, D. J. Adoption of Image-Driven Machine Learning for Microstructure Characterization and Materials Design: A Perspective. *JOM* **2021**, *73* (11), 3639−3657.

(27) Ede, J. M. Deep learning in electron microscopy. *Machine Learning: Sci. Technol.* **2021**, *2* (1), 011004.

(28) Melanthota, S. K.; Gopal, D.; Chakrabarti, S.; Kashyap, A. A.; Radhakrishnan, R.; Mazumder, N. Deep learning-based image processing in optical microscopy. *Biophys. Rev.* **2022**, *14*, 463−481.

(29) Jacobs, R. Deep learning object detection in materials science: Current state and future directions. *Comput. Mater. Sci.* **2022**, *211*, 111527.
(30) Treder, K. P.; Huang, C.; Kim, J. S.; Kirkland, A. I. Applications of deep learning in electron microscopy. *Microscopy* **2022**, *71* (Supplement_1), i100−i115.
(31) Zhang, L.; Shao, S. Image-based machine learning for materials science. *J. Appl. Phys.* **2022**, *132* (10), 100701.
(32) Botifoll, M.; Pinto-Huguet, I.; Arbiol, J. Machine learning in electron microscopy for advanced nanocharacterization: current developments, available tools and future outlook. *Nanoscale Horiz.* **2022**, *7* (12), 1427−1477.
(33) Kalinin, S. V.; Ghosh, A.; Vasudevan, R.; Ziatdinov, M. From atomically resolved imaging to generative and causal models. *Nat. Phys.* **2022**, *18* (10), 1152−1160.
(34) Unruh, D.; Kolluru, V. S. C.; Baskaran, A.; Chen, Y.; Chan, M. K. Theory+AI/ML for microscopy and spectroscopy: Challenges and opportunities. *MRS Bull.* **2022**, *47* (10), 1024−1035.
(35) Lyu, Z.; Yao, L.; Chen, W.; Kalutantirige, F. C.; Chen, Q. Electron Microscopy Studies of Soft Nanomaterials. *Chem. Rev.* **2023**, *123* (7), 4051−4145.
(36) Shen, D.; Wu, G.; Suk, H.-I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221.
(37) Kermany, D. S.; Goldbaum, M.; Cai, W.; Valentim, C. C.; Liang, H.; Baxter, S. L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.; et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **2018**, *172* (5), 1122−1131.
(38) Lu, S.; Montz, B.; Emrick, T.; Jayaraman, A. Semi-supervised machine learning workflow for analysis of nanowire morphologies from transmission electron microscopy images. *Digital Discovery* **2022**, *1* (6), 816−833.
(39) Liu, D.; Tan, Y.; Khoram, E.; Yu, Z. Training deep neural networks for the inverse design of nanophotonic structures. *ACS Photonics* **2018**, *5* (4), 1365−1369.
(40) So, S.; Mun, J.; Rho, J. Simultaneous inverse design of materials and structures via deep learning: demonstration of dipole resonance engineering using core-shell nanoparticles. *ACS Appl. Mater. Interfaces* **2019**, *11* (27), 24264−24268.
(41) Stein, H. S.; Guevarra, D.; Newhouse, P. F.; Soedarmadji, E.; Gregoire, J. M. Machine learning of optical properties of materials-predicting spectra from images and images from spectra. *Chem. Sci.* **2019**, *10* (1), 47−55.
(42) Zhou, Q.; Yong, B.; Lv, Q.; Shen, J.; Wang, X. Deep autoencoder for mass spectrometry feature learning and cancer detection. *IEEE Access* **2020**, *8*, 45156−45166.
(43) Banko, L.; Maffettone, P. M.; Naujoks, D.; Olds, D.; Ludwig, A. Deep learning for visualization and novelty detection in large X-ray diffraction datasets. *npj Comput. Mater.* **2021**, *7* (1), 104.
(44) Grossutti, M.; D'Amico, J.; Quintal, J.; MacFarlane, H.; Quirk, A.; Dutcher, J. R. Deep Learning and Infrared Spectroscopy: Representation Learning with a β-Variational Autoencoder. *J. Phys. Chem. Lett.* **2022**, *13* (25), 5787−5793.
(45) He, C.; Zhu, S.; Wu, X.; Zhou, J.; Chen, Y.; Qian, X.; Ye, J. Accurate Tumor Subtype Detection with Raman Spectroscopy via Variational Autoencoder and Machine Learning. *ACS Omega* **2022**, *7* (12), 10458−10468.
(46) Yaman, M. Y.; Kalinin, S. V.; Guye, K. N.; Ginger, D. S.; Ziatdinov, M. Learning and Predicting Photonic Responses of Plasmonic Nanoparticle Assemblies via Dual Variational Autoencoders. *Small* **2023**, *19* (25), 2205893.
(47) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **2018**, *4* (2), 268−276.
(48) Rajak, P.; Krishnamoorthy, A.; Nakano, A.; Vashishta, P.; Kalia, R. Structural phase transitions in a $MoWSe_2$ monolayer: Molecular dynamics simulations and variational autoencoder analysis. *Phys. Rev. B* **2019**, *100* (1), 014108.
(49) Christofi, E.; Chazirakis, A.; Chrysostomou, C.; Nicolaou, M. A.; Li, W.; Doxastakis, M.; Harmandaris, V. A. Deep convolutional neural networks for generating atomistic configurations of multi-component macromolecules from coarse-grained models. *J. Chem. Phys.* **2022**, *157* (18), 184903.
(50) Tung, C.-H.; Chang, S.-Y.; Chang, M.-C.; Carrillo, J.-M.; Sumpter, B. G.; Do, C.; Chen, W.-R. Inferring colloidal interaction from scattering by machine learning. *Carbon Trends* **2023**, *10*, 100252.
(51) Cang, R.; Li, H.; Yao, H.; Jiao, Y.; Ren, Y. Improving direct physical properties prediction of heterogeneous materials from imaging data via convolutional neural network and a morphology-aware generative model. *Comput. Mater. Sci.* **2018**, *150*, 212−221.
(52) Kim, Y.; Park, H. K.; Jung, J.; Asghari-Rad, P.; Lee, S.; Kim, J. Y.; Jung, H. G.; Kim, H. S. Exploration of optimal microstructure and mechanical properties in continuous microstructure space using a variational autoencoder. *Mater. Des.* **2021**, *202*, 109544.
(53) Yamashita, A.; Nagata, T.; Yagyu, S.; Asahi, T.; Chikyow, T. Direct feature extraction from two-dimensional X-ray diffraction images of semiconductor thin films for fabrication analysis. *Sci. Technol. Adv. Mater.: Methods* **2022**, *2* (1), 23−37.
(54) Bridger, A.; David, W. I.; Wood, T. J.; Danaie, M.; Butler, K. T. Versatile domain mapping of scanning electron nanobeam diffraction datasets utilising variational autoencoders. *npj Comput. Mater.* **2023**, *9* (1), 14.
(55) Chen, W.; Ferguson, A. L. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *J. Comput. Chem.* **2018**, *39* (25), 2079−2102.
(56) Chen, W.; Tan, A. R.; Ferguson, A. L. Collective variable discovery and enhanced sampling using autoencoders: Innovations in network architecture and error function design. *J. Chem. Phys.* **2018**, *149* (7), 072312.
(57) Yin, J.; Pei, Z.; Gao, M. C. Neural network-based order parameter for phase transitions and its applications in high-entropy alloys. *Nat. Comput. Sci.* **2021**, *1* (10), 686−693.
(58) O'Leary, J.; Mao, R.; Pretti, E. J.; Paulson, J. A.; Mittal, J.; Mesbah, A. Deep learning for characterizing the self-assembly of three-dimensional colloidal systems. *Soft Matter* **2021**, *17* (4), 989−999.
(59) Mao, R.; O'Leary, J.; Mesbah, A.; Mittal, J. A Deep Learning Framework Discovers Compositional Order and Self-Assembly Pathways in Binary Colloidal Mixtures. *JACS Au* **2022**, *2* (8), 1818−1828.
(60) Kingma, D. P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv (Statistics.Machine Learning)*, December 10, 2022, 1312.6114, ver. 11. https://arxiv.org/abs/1312.6114 (accessed 2023-04-11).
(61) Fasolka, M. J.; Mayes, A. M. Block copolymer thin films: Physics and applications. *Annu. Rev. Mater. Res.* **2001**, *31* (1), 323−355.
(62) Bockstaller, M. R.; Mickiewicz, R. A.; Thomas, E. L. Block copolymer nanocomposites: perspectives for tailored functional materials. *Adv. Mater.* **2005**, *17* (11), 1331−1349.
(63) Orilall, M. C.; Wiesner, U. Block copolymer based composition and morphology control in nanostructured hybrid materials for energy conversion and storage: solar cells, batteries, and fuel cells. *Chem. Soc. Rev.* **2011**, *40* (2), 520−535.
(64) Doerk, G. S.; Yager, K. G. Beyond native block copolymer morphologies. *Mol. Syst. Des. Eng.* **2017**, *2* (5), 518−538.
(65) Hajduk, D. A.; Harper, P. E.; Gruner, S. M.; Honeker, C. C.; Kim, G.; Thomas, E. L.; Fetters, L. J. The gyroid: a new equilibrium morphology in weakly segregated diblock copolymers. *Macromolecules* **1994**, *27* (15), 4063−4075.
(66) Khandpur, A. K.; Foerster, S.; Bates, F. S.; Hamley, I. W.; Ryan, A. J.; Bras, W.; Almdal, K.; Mortensen, K. Polyisoprene-polystyrene diblock copolymer phase diagram near the order-disorder transition. *Macromolecules* **1995**, *28* (26), 8796−8806.
(67) Cochran, E. W.; Garcia-Cervera, C. J.; Fredrickson, G. H. Stability of the gyroid phase in diblock copolymers at strong segregation. *Macromolecules* **2006**, *39* (7), 2449−2451.

https://doi.org/10.1021/jacsau.3c00275
JACS Au 2023, 3, 2510−2521

(68) Pedersen, J. S.; Gerstenberg, M. C. Scattering form factor of block copolymer micelles. *Macromolecules* **1996**, *29* (4), 1363−1365.
(69) Alina, G.; Attala, Z.; Bakker, J.; Bourne, R.; Bouwman, W.; Breßler, I.; Butler, P.; Caddy-Jones, I.; Campbell, K.; Cho, J.-H.; et al. *SasView for Small Angle Scattering Analysis*. www.sasview.org (accessed 2023-05-23).
(70) Beltran-Villegas, D.; Heil, C.; Jayaraman, A.; Wessels, G. M.; Wu, Z.; Ye, Z. *CREASE Overview*, 2023. https://crease-ga.readthedocs.io/en/latest/index.html (accessed 2023-05-23).
(71) Doerk, G. S.; Stein, A.; Bae, S.; Noack, M. M.; Fukuto, M.; Yager, K. G. Autonomous discovery of emergent morphologies in directed self-assembly of block copolymer blends. *Sci. Adv.* **2023**, *9* (2), eadd3687.
(72) Yager, K. G.; Lai, E.; Black, C. T. Self-assembled phases of block copolymer blend thin films. *ACS Nano* **2014**, *8* (10), 10582−10588.
(73) Blaiszik, B.; Chard, K.; Pruyne, J.; Ananthakrishnan, R.; Tuecke, S.; Foster, I. The materials data facility: data services to advance materials science research. *JOM* **2016**, *68* (8), 2045−2052.
(74) Blaiszik, B.; Ward, L.; Schwarting, M.; Gaff, J.; Chard, R.; Pike, D.; Chard, K.; Foster, I. A data ecosystem to support machine learning in materials science. *MRS Commun.* **2019**, *9* (4), 1125−1133.
(75) Bradley, D.; Roth, G. Adaptive Thresholding using the Integral Image. *J. Graphics Tools* **2007**, *12* (2), 13−21.
(76) Torquato, S.; Stell, G. Microstructure of two-phase random media. I. The n-point probability functions. *J. Chem. Phys.* **1982**, *77* (4), 2071−2077.
(77) Lu, B.; Torquato, S. Lineal-path function for random heterogeneous materials. *Phys. Rev. A* **1992**, *45* (2), 922.
(78) Santos, J. E.; Xu, D.; Jo, H.; Landry, C. J.; Prodanović, M.; Pyrcz, M. J. PoreFlow-Net: A 3D convolutional neural network to predict fluid flow through porous media. *Adv. Water Res.* **2020**, *138*, 103539.
(79) Huang, Y.; Xiang, Z.; Qian, M. Deep-learning-based porous media microstructure quantitative characterization and reconstruction method. *Phys. Rev. E* **2022**, *105* (1), 015308.
(80) Gostick, J. T.; Khan, Z. A.; Tranter, T. G.; Kok, M. D.; Agnaou, M.; Sadeghi, M.; Jervis, R. PoreSpy: A python toolkit for quantitative analysis of porous media images. *J. Open Source Software* **2019**, *4* (37), 1296.
(81) Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Lecture Notes in Computer Science, Vol. *9351*; Springer, 2015; pp 234−241..

https://doi.org/10.1021/jacsau.3c00275
JACS Au 2023, 3, 2510−2521