Finding simplicity: unsupervised discovery of features, patterns, and order parameters via shift-invariant variational autoencoders

Maxim Ziatdinov$^{1,2,*}$, Chun Yin (Tommy) Wong$^{3,**}$ and Sergei V Kalinin$^{4,**}$

$^1$ Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America
$^2$ Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America
$^3$ Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN 37996, United States of America
$^4$ Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN 37996, United States of America

Authors to whom any correspondence should be addressed.

E-mail: maxim.ziatdinov@gmail.com, cwong13@vols.utk.edu and sergei2@utk.edu

Keywords: variational autoencoder, scanning tunneling microscopy, scanning transmission electron microscopy, unsupervised learning, latent representation

Abstract
Recent advances in scanning tunneling and transmission electron microscopies (STM and STEM) have allowed routine generation of large volumes of imaging data containing information on the structure and functionality of materials. The experimental data sets contain signatures of long-range phenomena such as physical order parameter fields, polarization, and strain gradients in STEM, or standing electronic waves and carrier-mediated exchange interactions in STM, all superimposed onto scanning system distortions and gradual changes of contrast due to drift and/or mis-tilt effects. Correspondingly, while the human eye can readily identify certain patterns in the images such as lattice periodicities, repeating structural elements, or microstructures, their automatic extraction and classification are highly non-trivial and universal pathways to accomplish such analyses are absent. We pose that the most distinctive elements of the patterns observed in STM and (S)TEM images are similarity and (almost-) periodicity, behaviors stemming directly from the parsimony of elementary atomic structures, superimposed on the gradual changes reflective of order parameter distributions. However, the discovery of these elements via global Fourier methods is non-trivial due to variability and lack of ideal discrete translation symmetry. To address this problem, we explore the shift-invariant variational autoencoders (shift-VAEs) that allow disentangling characteristic repeating features in the images, their variations, and shifts that inevitably occur when randomly sampling the image space. Shift-VAEs balance the uncertainty in the position of the object of interest with the uncertainty in shape reconstruction. This approach is illustrated for model 1D data, and further extended to synthetic and experimental STM and STEM 2D data. We further introduce an approach for training shift-VAEs that allows finding the latent variables that comport to known physical behavior. In this specific case, the condition is that the latent variable maps should be smooth on the length scale of the atomic lattice (as expected for physical order parameters), but other conditions can be imposed. The opportunities and limitations of the shift VAE analysis for pattern discovery are elucidated.

Notice: This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC0500OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for the United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

© 2023 The Author(s). Published by IOP Publishing Ltd
The progress in atomically resolved scanning probe and electron microscopy over the last three decades has opened the pathway to high veracity imaging of atomic structures and electronic phenomena in a broad variety of materials. By now, high-resolution (scanning) transmission electron microscopy (STEM) has emerged as a primary tool for studies of bulk and 2D materials, capable of visualizing the atomic structure of defects, interfaces, and topological defects associated with chemical and ferroic ordering [1–6]. Similarly, scanning tunneling microscopy (STM) and non-contact atomic force microscopy have become primary tools for exploring complex electronic orders in charge density wave and superconducting systems [7–11].

The emergence of large volumes of microscopic data at or near atomic resolution necessitates the development of methods capable of extracting relevant information from large volumes of data. For hyperspectral data sets such as tunneling spectroscopy in STM, force–distance curves in atomic force microscopy (AFM), and electron energy loss spectroscopy in STEM, progress was achieved with the introduction of multivariate statistical methods such as principal component analysis [12, 13] and more complex linear dimensionality reduction methods [14–16], and later simple and variational autoencoders (VAEs) [17]. However, these methods generally apply to 1 (or higher D) spectral data, while correlations in the image plane remain unused and are explored only via human inspection of the corresponding loading/abundance maps.

The advances in machine learning (ML) for image analysis over the last decade have stimulated interest in applications of these methods in physical imaging. The supervised ML methods have been broadly incorporated in the analysis of electron and scanning probe microscopy data sets, for example towards atomic and molecular identification [18–21]. It has also been applied in computational research and modeling of crystal structures [22, 23]. Similar approaches were broadly used for a variety of applications including the search for exotic particles from the Large Hadron Collider data [24], analysis of satellite images [25], and cancer detection [26]. However, supervised ML methods explicitly rely on the availability of the labeled data. These typically require human labeling or in special cases using theoretical models.

Over the last year, there has been a progressively growing awareness both in the ML community and the domain sciences that applications of supervised ML methods in experimental sciences are associated with a number of unique challenges. Indeed, classical ML applications deal with large data volumes, the proverbial ‘big data’, that sufficiently sample the full distribution of the data. For example, the handwritten digits in the MNIST data set or cat images in ImageNet represent a significant subset of all possible writing styles or colorations and poses that cats can adopt, and hence applications to an unknown digit or cat image are in-distribution problems. Comparatively, experimental data sets are sensitive to exact image acquisition parameters and often lead to out-of-distribution effects, when the supervised algorithm trained for specific conditions will perform poorly for different data acquisition conditions. These issues and some strategies to mitigate these are analyzed in detail in [27].

These considerations in turn necessitate the development of unsupervised methods for the analysis of the 2D and hyperspectral images. However, unsupervised methods implicitly rely on inferential biases encoded in the architecture of neural networks and loss functions, as well as the selection (or engineering) of the feature set [28]. Classical convolutional neural networks (CNNs) have been used for both supervised learning [17–20, 22, 23] in pattern recognition and unsupervised representation learning [29]. For the neural network architectures, it is generally recognized that CNNs are translationally-equivariant, i.e. during training they can translate knowledge learned in one location in an image to other locations. At the same time, their response shifts with the position of an object, that is, they do not have a positional invariance [30]. As a result, CNNs perform poorly when tasked with learning a representation of arbitrarily shifted objects [30, 31]. The addition of max-pooling layers [32] can partially address this issue if the shifts are (very) small, but not for moderate and large shifts.

An equally important role is played by feature selection. In the absence of any prior data, the images can be represented by a collection of patches centered on the uniform grid that (along with the original position) for a feature set for unsupervised learning. In many cases, imaging data contains physically-defined locations that offer a natural way to define features, e.g. atomic positions in STEM data or the center of mass of particles. In these cases, the image patches can be selected centered on these features, and by construction, the application of unsupervised ML methods will provide information on local chemical and physical phenomena [33–40]. Here, one of the major limitations is the potential presence of multiple versions of the same object with dissimilar rotations within the image plane, either due to disorder in materials or distortions during the scanning process. The recent introduction of the spatial VAE has mitigated this issue [41].

However, the availability of physics-defined reference points is not universal to all imaging modes and is directly linked to the signal-formation mechanisms. For example, STEM is preponderantly sensitive to atomic nuclei, and hence contrast maxima provide reliable information on atomic positions. At the same time, STM is sensitive to the electronic density of states. Hence the observed contrast can be expected to have
the periodicity of the underlying lattice; however, the observed features can have an arbitrary registry to the lattice depending on local Fermi level position, the proximity of defects, and other factors. Hence, of interest is the separation of the building blocks of such images and the remaining factors of variation across the image space that are changing smoothly almost everywhere on the length scale of atomic units and can correspond to physical order parameter fields or microscope distortions.

Here we explore a shift-invariant VAE (shift-VAE) concept as a universal method for the analysis of such data sets. We pose that the most distinctive elements of the patterns observed in STM and (S)TEM images are similarity and (almost-) periodicity, behaviors stemming directly from the parsimony of elementary atomic structures. This allows for the extension of representation learning—realized here via the shift-VAE—to balance the uncertainty in the position of the object of interest with the uncertainty in shape reconstruction. This approach is illustrated for the model 1D data and further extended to synthetic and experimental STM and STEM 2D data. The opportunities and limitations of the representation learning analysis for pattern discovery and order parameter field mapping are elucidated.

In representation learning, one assumes that complicated real-world observations originate from a small number of the explanatory factors of variation behind the data [42]. A VAE [43, 44] is a powerful latent variable framework to infer those (latent) factors given the data and some prior knowledge. The latter includes a fixed prior distribution over a low-dimensional latent space and, optionally, a small number of labeled data samples [45]. Practically, however, it is usually also necessary to fine-tune the VAE’s architecture as well as its loss objective for different problems [28]. A VAE consists of the stochastic inference (encoder) and generative (decoder) models parameterized by deep neural networks. The decoder links the latent space to the data space via the likelihood and generative (decoder) models parameterized by deep neural networks. The encoder maps the data space to the latent space via the posterior distribution \( q(z|x) \) where \( f_{\varphi} \) is a different neural network with parameters \( \varphi \). This formulation allows turning the VAE approach into an optimization problem where we want to learn the parameters of encoder and decoder neural networks. The loss objective, in this case, is a combination of data reconstruction error and the Kullback–Leibler (KL) divergence term between the encoded \( q(z|x) \) and prior \( p(z) \) distributions.

The presence of multiple translated and/or rotated instances of the same objects typically leads to the suboptimal disentanglement of the true physical factors of variation in VAE. Indeed, the Galilean principle [46] asserts that mechanical equations of motion are invariant under the \( r' = r + \Delta r \) transformation, which is not accounted for by the classical VAE. To incorporate knowledge of basic mechanics into the VAE architecture, we partition our latent space into conventional VAE latent variables and one or two ‘special’ latent variables associated with the variations in position. The latter is used to shift the 2D grid coordinates of the input data by \( k\Delta r \) where \( 0 < k \leq 1 \) incorporates our ‘belief’ about a degree of disorder in the system. We refer to this specific modification of VAE as shift-VAE (see schematic in figure 1). The loss objective is defined as:

\[
\mathcal{L} = \mathbb{E}_{x \sim p_{\text{data}}} \mathcal{L}_{\text{RE}} + \beta(t) \mathcal{L}_{\text{KL}},
\]

where \( \mathcal{L}_{\text{RE}} \) is a reconstruction error, \( \mathcal{L}_{\text{KL}} \) corresponds to the KL divergence term and \( \beta > 0 \) is usually introduced to encourage a better disentanglement of the latent representations [47].

To get insight into the principles and operation of the shift-VAE, we first illustrate this approach for the one-dimensional synthetic data. We define the family of \( N = 5000 \) sine wave segments with the defined periodicity and varying offsets as \( y(x) = \sin((x - \mu)/\sigma) + \sigma_{\text{noise}} \) defined on \( x \in [-10, 10] \). The offsets are uniformly distributed on \( \mu \in [-3, 3] \), whereas periodicities are chosen to have (small) discrete number of values. This choice is made both to explore the relationship between the factors of variability in data and VAE performance, and to approximate an experimental scenario when only a few possible wavelengths are present in a single image. White noise of amplitude \( \sigma_{0i} \) or amplitude uniformly distributed on \( \sigma_{\text{noise}} \in [0, 0.1] \) interval, can be added. Examples of the data set are available through the accompanying Google Colab notebook.

The analysis of the data using a regular VAE is shown in figures 2(a) and (b) for several distributions of periodicity. In this case, the encoded data shows a set of well-formed 1D manifolds in the latent space, with the phase changing uniformly across the manifold. Note that the manifolds corresponding to different periods are not necessarily continuous and can be split into two or more. These behaviors can be understood as the result of the countervailing tendencies of VAEs to represent the data in the compact form and at the same time from the fact that manifolds for different periodicities cannot cross (since cross-point will be unphysical). These trends are even more pronounced for the larger number of periodicities, as shown in figure 2(b). It is also important to note that these patterns are closely related to the sampling of parameter space, i.e. a large number of possible offsets and a small number of periodicities (but this, in turn, is chosen to comport to plausible experimental scenarios).
Figure 1. Schematic of the shift-VAE. The encoder maps data $x$ of dimensionality $d$ ($d = 1, 2$) into the mean ($\mu$) and standard deviation ($\sigma$) of the normal distribution from which a latent variable $z$ with $n \times (l + d)$ dimensionality is sampled ($l$ is the dimensionality of the conventional VAE latent space). We take the first two columns of the $z$ vector ($\Delta r := z[1, : d]$) and use them to shift our coordinate grid by $k \Delta r$ where $0 < k \leq 1$ is fixed before the VAE training. The transformed grid is then concatenated with the remaining latent variables ($z := z[1, : d]$) and passed to a decoder [41]. Finally, we score the observed data against the Bernoulli ($\text{Ber}$) likelihood parameterized by the decoder output.

Note that despite the fact that for a simple VAE the latent variables do not allow direct interpretation, they nonetheless allow exploring the characteristic features of data. Here, the postulated premise is that real materials will have only a small set of atomic spacings as imposed by preferred bond lengths. In the presence of noise or a more complex structure, these manifolds will broaden. The corresponding projections of the latent space representations to data space are shown in figures 2(d) and (e), and visualize data decoded from the latent points distributed over the uniform spatial grid. The reconstructed objects clearly illustrate sine waves, and these behaviors are further analyzed later.

In comparison, shown in figure 2(c) is the latent space for shift-VAE. Here, one of the latent variables is an offset, and the remaining two are the classical continuous latent variables. Note that in this case the manifolds corresponding to different periodicities are clearly separated. Furthermore, the latent space is clearly quasi-collapsed, i.e. data points are distributed only along one latent direction. While typically perceived to be a problem for VAEs, we pose that this behavior in fact allows inferring the physics of the system, namely the number of the physical factors of variability. In this case, the synthetic data set has two factors of variability—phase and periodicity, with the phase separated as a shift. Correspondingly, in this case, periodicity becomes associated with one of the latent variables, and the second latent variable collapses, in agreement with this conjecture.

We further explored the structure of the latent space for the VAE and shift-VAE, namely the relationship between the latent parameters and physical descriptors. Specifically, we decoded the signal generated with parameters $\mu \in [-3, 3], \sigma \in [1, 2, 2.0, 2.2]$ from each point on the latent space grid and fit it to the sine wave. The fitting function is defined as $y(x) = \hat{\delta} \sin ((x - \hat{\mu}) / \hat{\sigma}) + \hat{\sigma}_{\text{noise}}$, where $\hat{\delta}$ is the amplitude and the hat indicates fitted parameters as opposed to parameters used to generate the data.
Figure 2. Latent space distributions formed by the data encoded with a regular VAE for (a) $\sigma \in [1, 1.6, 0.2]$ and (b) $\sigma \in [1, 2.4, 0.2]$. (c) Latent space distribution formed by the data encoded with the shift-V AE for $\sigma \in [1, 2.4, 0.2]$. (d)–(f) Corresponding learned latent space representations projected to the data space. Here, the distribution of periodicities is defined as $\sigma \in [\sigma_{\text{min}},\sigma_{\text{max}}]$. The fit parameters, $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\delta}$, at each pair of latent parameters are shown in figures 3(a)–(c) for V AE and (d)–(f) for shift-V AE. Note that for simple V AE the parameters in latent space are distributed following complex patterns, as can be expected from original data distributions shown in figures 2(a) and (b). At the same time, figures 3(d)–(f) illustrate the efficacy of the shift-V AE on separating sine curves with different fit parameters. The efficacy, however, drops sharply with increasing variances in the fit parameters. In such a manner, we establish the relationship between pairs of encoding latent variables ($z_1, z_2$) and the parameters $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\delta}$. We can also plot the fit error, defining the discrepancy between the data and the functional form of the ground truth. Note that for shift-V AE the expected error for the offset variable should be zero by construct, and hence deviation from zero provides a measure of the disentanglement of the shift variable, as will be explored below.

Finally, we note that the convergence of the shift-V AE strongly depends on the distribution of the initial data. For small and moderate phase offsets, the recovered shift value is equal to the ground truth offset. For larger offsets, the behavior starts to deviate from the ground truth value. Furthermore, in some cases the training process leads to the emergence of the metastable minima, i.e. the manifold for the single periodicity will be split in two by a band of minima in $z_2$ as shown in figures 3(d)–(f). The observed minima are metastable because as observed in figure 3(g), for high $\hat{\mu}$ values, any minor perturbations could significantly decrease the fidelity of the fitted parameter to the ground truth as measured by the $R^2$ value, corresponding to the collapse of the latent representations.

To explore these behaviors systematically, shown in figures 3(g) and (h) are the $R^2$ and the ratio $\mu : \hat{\mu}$ for pairs of $\mu$ spread and $\sigma_{\text{noise}}$ values. Here, $\mu$, $\sigma$, and $\delta$ are varied independently from $\mu \in [-6, 6]$, $\sigma \in [1.8, 12]$, and $\delta \in [1, 12]$. The error in the sine fit is then evaluated by first performing linear regression on the latent variable against the fitted parameter and evaluating both the $R^2$ coefficient for the quality of the fit, and the ratio between the latent variable and the fitted parameter. Here, a ratio greater than 1 means that the latent variable overpredicts the fit parameter of the sine wave, and a value less than one means that the latent variable underpredicts the fit parameter. Worth noting is the sharp decrease in the quality of the fit for $\mu > 5.6$ and $\sigma_{\text{noise}} > 1$ in figure 3(g), meaning that outside the yellow region, the shift-V AE disentanglement of the shift variable fails. Similarly, figure 3(h) reveals that there is a gradual underprediction of the latent variable before a sharp overprediction of the latent variable, resulting in a high instability at the boundary. We further note that this behavior can be controlled through the choice of the priors on the variational distributions and the loss function, and an approach for this will be illustrated below.
Figure 3. The sinusoidal fit parameters in the VAE latent space. Shown are maps for (a) fitted shift, \( \hat{\mu} \), (b) fitted period, \( \hat{\sigma} \), and (c) fitted amplitude, \( \hat{\delta} \), respectively. Fitted parameter maps in shift-VAE latent space for (d) fitted shift, \( \hat{\mu} \), (e) fitted period, \( \hat{\sigma} \), and (f) fitted amplitude, \( \hat{\delta} \). Compared to the VAE latent space, the shift-VAE latent space shows several distinct linear regions corresponding to the transition of the fit parameters. Figures (g) and (h) show the error in the fit in terms of \( R^2 \) and the ratio \( \mu : \hat{\mu} \), delineating the region where shift-VAE latent shift coincides with ground truth value.

We further extended the shift-VAE approach to a 2D case. In this case, we introduce two special latent variables designated to encode the information about the offsets, giving rise to \( 2 + l \) latent space. For the convenience of representation, we choose \( l = 2 \). As a synthetic data set, we use a simple square lattice with ideal translational symmetry and a 2D Gaussian centered at each lattice site as shown in figure 4(a). To create image patches, the sampling points are chosen with uniformly distributed random offsets with respect to the ideal lattice position, as shown in figure 4(b). Here, if the distribution is very narrow, this implies that all the patches are centered at atomic units. If the distribution is broad (or sampling points are chosen randomly), the patches do not contain information on atomic positions.

The simple VAE analysis of this created data set with an angle of rotation between \(-0.035\) rad and \(0.035\) rad, and a displacement of atomic positions \([0,0.1]\) is illustrated in figures 4(c) and (d). Here, the projected latent manifold of the VAE is populated by the image patches with different offsets. The corresponding encoding of the data into the latent space shown in figure 3(d) has a very complex structure formed by the multiple linear manifolds. The emergence of this structure is attributed to the digitization effects, i.e. a relatively low number of pixels in the image patches. However, we note that the conditions here are chosen to be close to experimental, and in the subsequent discussion we will see that similar behavior will manifest in the experimental data. The color represents the displacement of the atomic position.

In comparison, the latent space of the shift-VAE projected to the data space appears to be populated by identical elements (figure 4(e)). Correspondingly, the associated latent distribution is highly localized and both continuous latent variables are collapsed (figure 4(f)). Similar to the example of 1D data set, we attribute this behavior to the fact that offsets were the only physical factors of variability in the original data.
set. Here, the offsets separated as two special latent variables are the only two factors of variation and it is unsurprising that the remaining variability is absent and latent space is doubly degenerate.

With these analyses in hand, we extended this approach to the experimental data sets. While conventional VAE has been applied to experimental STEM data before [29], the capability of shift-VAE in asserting translational and rotational invariances has yet to be demonstrated. As models, here we use the STM images of graphene (figure 5(a)), and STEM images of BiFeO$_3$ (figure 5(b)) and NiO-LSMO heterostructure (figure 5(c)). These datasets have been used in prior publications [27, 48, 49], which give the details on the nature of the system and imaging conditions. Here, they are used as representative examples of systems with complex spatially-varying patterns with atomic resolution. The selection of these three systems is predicated by the nature of anticipated behaviors, namely perturbation of the electronic structure of graphene on a large number of individual defects, presence of two phases, ferroic domains, and chemical variability in BiFeO$_3$, and two-phase coexistence in the NiO-LSMO composite.

Similar to the synthetic data, the images are transformed into a collection of patches sampled over the square grid. The window size and the grid sampling can be varied to match the specific physics of the image.
Here we have found that optimal performance can be obtained for patch size incorporating 2–3 unit cells, with the grid step being $1/2$ of the patch size. Note that the grid spacing in this case is chosen randomly and does not contain any information on atomic positions. The provided Google Colab notebook allows readers to experiment with these parameters or analyze their own data. However, it is important to note that the positions of the atomic units (as used previously in $[50, 51]$) in this case are not available and the choice of the rectangular grid represents a lack of prior knowledge or physical inferential biases. We also note that for graphene data we found that each patch should be normalized; otherwise, the relative intensity changes become the primary factor of variation in data.

The VAE analysis for chosen systems is illustrated in figure 6. In all cases, we show the (i) learned latent manifolds decoded to the data space, (ii) encoded latent distributions, and (iii) spatial maps of the encoded latent variables. Note that compared to the idealized model system in figure 4, these images represent both partially known physics of these materials, and also contain the non-idealities of the real imaging system.

For graphene, the learned latent space representation (figure 6(a-i)) contains elements resembling the ideal hexagonal graphene lattice, and elements resembling waves in one of the chosen lattice directions. The distribution of these in the latent space is fairly complex, as is typical for the system having rotational degeneracies (which in this case are not captured by the latent variable). The corresponding encoded distribution (figure 6(a-ii)) is relatively simple and contains a small number of crisscrossing features. The spatial latent variable maps clearly separate the defect and non-defect regions and are dominated by the

---

**Figure 6.** The regular VAE analysis of the experimental data: (i) the learned latent manifolds projected to the data space, (ii) the encoded latent distributions, where the color bar represents the kernel density of the latent distributions, and (iii) the latent variables, $z_1, z_2$, in the image spaces for (a) graphene, (b) BiFeO$_3$, and (c) NiO-LSMO.
high-periodicity features formed by the beating between the sampling square lattice and intrinsic material lattice. For BiFeO₃, the learned latent space is populated by the distorted versions of the perovskite unit cell (figure 6(b-i)), whereas for NiO-LSMO system careful examination illustrate presence of A- and B-site centered blocks and rocksalt blocks (figure 6(c-i)). The corresponding encoded latent distributions (figures 6(b-ii) and (c-ii)) are very complex, resembling that in figure 4(d). The corresponding spatial maps are again dominated by the beating effects.

In comparison, the corresponding shift-VAE analyses are shown in figure 7. Here, shown are (i) the learned latent manifolds projected to the data space and (ii) the encoded latent distributions as well as the spatial maps corresponding to (iii) two conventional latent variables and (iv) two offset latent variables. Note that in this case the learned latent manifolds are formed by identical or weakly-changing blocks, whereas the encoded latent distributions are particularly simple and formed by the small number of modes. By inspection of the spatial maps, the physical order parameters can be identified. The spatial maps of graphene (figure 7(a)) exhibit regions of differing degrees of electronic disorder, however further analysis is required to understand the exact physical properties of the different regions. The spatial maps of the BiFeO₃ sample (figure 7(b)) are associated with the substrate and ferroic variants in the BiFeO₃ system. The micrograph does not show visibly distinct domains, but they are identified by visualizing the latent variables on the real image space. The spatial maps of the NiO-LSMO sample (figure 7(c)) clearly identify the two phases as observed visually on the micrograph, demonstrating the efficacy of shift-VAE.

In the corresponding spatial maps, the latent shifts show the beating pattern as can be expected, whereas the latent variables show smooth maps. Here, we note that this is the required behaviors, since typically order parameters in physical systems are postulated to change smoothly on the length scale of the unit cell.

Finally, we note that for a broad offset distribution, the shift-VAE can converge to multiple minima, and under certain conditions the high-frequency periodic features can 'leak' from the offset latent variables to conventional latent variables. We found empirically that this process can be controlled by the relative weights of the reconstruction and KL terms expressed via a scale factor \( \beta \) in equation (1). We further systematized this analysis by exploring the effect of this parameter as a function of time to control the shift-VAE training.

We note that to find an optimal KL scale factor for obtaining physically relevant latent representations, one must introduce a 'physicality' parameter defining how the result should 'look like'. Here, we take advantage of the fact that an order parameter field has to be smooth on the length scale of atomic units. This means that latent variables change smoothly over the image, whereas latent offset images show periodicity corresponding to the Moire pattern between the sampling lattice and intrinsic atomic lattice of materials.

To implement this, a grid search is carried out for KL scale factors ranging from 0.1 to 10 for initial and final KL values, respectively. The Fourier transformation is computed for the resultant latent variables, \( z_1 \) and \( z_2 \), in image spaces. The ratio of central peak intensity to the total image intensity yields the measure of smoothness of the data since the periodic features yield the peaks at finite values of the reciprocal lattice vectors. The greater of the intensities computed from the two latent variables is used to generate the heat map as a measure of the performance of the shift-VAE.

The resultant dependencies are shown in figure 8. Here, the final value of the KL scale factor affects the resultant latent maps only weakly, whereas the initial value has a strong effect on the convergence of the shift-VAE. For small initial values of the KL scale factor, the resultant maps are smooth, i.e. comport to the behavior expected to physical order parameters. For larger KL values one of the maps starts to exhibit the periodic features due to the beating between sampling and the atomic lattices, and for even large parameter values both maps contain these high-frequency features. It is important to note that within these individual regions the maps are generally similar, i.e. once the right region of the hyperparameter space is identified the answers (for the 'smoothness' parameter) do not depend on the exact KL value.

This behavior can be rationalized from an information theory point of view. The larger penalties on the KL term result in a significantly restricted capacity of the latent bottleneck. In this type of scenario, the optimal behavior of (shift-)VAE is to encode only the positional information since it can provide the most gains for data log-likelihood (reconstruction term in equation (1)) compared to other factors of variation. This forces the shift-VAE to start encoding the information about shifts into a conventional latent variable instead of the special offset latent variable leading to the aforementioned leakage. By starting with a smaller KL scale factor and gradually increasing it, we encourage our model to encode other (than shift) factors of variation into the conventional latent variables and force the learned latent factors to be more independent, while preserving a high-quality data reconstruction (not shown).

To summarize, we introduced a shift-VAE approach for the discovery of repeated features and patterns in images. This approach can be used for physics discovery in atomically resolved scanning transmission electron microscopy and STM data and is made possible by the parsimony of elementary atomic structures. Importantly, it allows separating the individual building blocks present in a small number of discrete classes.
from the relevant order parameter fields. For the latter, we assume that they (almost everywhere) change slowly on the length scale of the atomic lattice.

As an example of a use case, it would be possible to perform real-time analysis of STEM data of ferroelectric materials and the associated ferroelectric domains. Typically, the STEM images are analyzed post-acquisition, and any subsequent analysis will require locating the atomic positions first. We have previously demonstrated that it is possible to perform VAE analysis post-acquisition without finding the
atomic positions but the analysis is not robust [52]. By utilizing the aforementioned sliding window algorithm coupled with shift-VAE, we have demonstrated that a robust analysis can be carried out post-acquisition. This can then be extended to the real-time analysis of STEM data.

We further demonstrated that the shift-VAE training can be guided through the optimization of the KL scale factors, allowing to incorporate the inferential biases into the system. We note that in the online settings (i.e. during the actual STM/STEM experiments) the grid search strategy used here may not be the optimal one. In the future, the utilization of the Bayesian optimization techniques [53, 54] may lead to better sampling efficiency and faster convergence to a physically-meaningful latent space.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/tommycwong/shift-VAE-atomic [55].

Acknowledgments

This work was supported in part (S.V.K.) by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences as part of the Energy Frontier Research Centers program: CSSAS-The Center for the Science of Synthesis Across Scales under award number DE-SC0019288; and (M.Z. and C.Y.T.W.) by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, MLExchange Project, award number 107514.

ORCID iDs

Maxim Ziatdinov 🌐 https://orcid.org/0000-0003-2570-4592
Chun Yin (Tommy) Wong 🌐 https://orcid.org/0000-0001-8625-8026
References

[1] Pennycook S J and Nellist P D 2011 Scanning Transmission Electron Microscopy: Imaging and Analysis (Springer)
[2] Oxley M P, Lupini A R and Pennycook S J 2017 Ultra-high resolution electron microscopy Rep. Prog. Phys. 80 64
[3] Pennycook S J 2017 The impact of STEM aberration correction on materials science Ultramicroscopy 180 22–33
[4] Das S et al 2019 Observation of room-temperature polar skyrmions Nature 568 368
[5] Delfyett N, Krivanek O L, Nellist P D, Batson P E and Lupini A R 2001 Progress in aberration-corrected scanning transmission electron microscopy J. Electron. Microsc. 50 177–86
[6] Taheri M I, et al 2016 Current status and future directions for in situ transmission electron microscopy Ultramicroscopy 170 86–95
[7] Pan S H, Hudson E W, Lang K M, Eisaki H, Uchida S and Davis J C 2000 Imaging the effects of individual zinc impurity atoms on superconductivity in Bi2Sr2CaCu2O8+δ Nature 403 746–50
[8] Heller E J, Crommie M F, Lutz C P and Eigler D M 1994 Scattering and absorption of surface electron waves in quantum corrals Nature 369 464–6
[9] Roushan P, Seo J, Parker C V, Hor Y S, Hsieh D, Qian D, Richardson A, Hasan M Z, Cava R J and Yazdani A 2009 Topological surface states protected from backscattering by chiral spin texture Nature 460 1106–11164
[10] Sugimoto Y, Pou P, Abe M, Jelinek P, Perez R, Morita S and Custance O 2007 Chemical identification of individual surface atoms by atomic force microscopy Nature 446 64–67
[11] Okamoto K, Sugawara Y and Morita S 2003 The imaging mechanism of atomic-scale Kelvin probe force microscopy and its application to plasmonic atomic-scale force mapping Japan. J. Appl. Phys. 41 7216–8
[12] Bosman M, Watanabe M, Alexander D T L and Keast V J 2006 Mapping chemical and bonding information using multivariate analysis of electron energy-loss spectrum images Ultramicroscopy 106 1024–32
[13] Jesse S and Kalinin S V 2009 Principal component and spatial correlation analysis of spectroscopic-imaging data in scanning probe microscopy Nanotechnology 20 085714
[14] Dobigeon N and Brun N 2012 Spectral mixture analysis of EELS spectrum-images Ultramicroscopy 120 25–34
[15] Halimi A, Dobigeon N and Tourneret J Y 2015 Unsupervised unmixing of hyperspectral images accounting for end member variability IEEE Trans. Image Process. 24 4900–17
[16] Kannan R, Ievlev A V, Laamait N, Ziatdinov M N, Vasudevan R K, Jesse S and Kalinin S V 2018 Deep data analysis via physically constrained linear unmixing: universal framework, domain examples, and a community-wide platform Adv. Struct. Chem. Imaging 4 20
[17] Agar J C et al 2019 Revealing ferroelectric switching character using deep recurrent neural networks Nat. Commun. 10 4809
[18] Ziatdinov M, Maksov A and Kalinin S V 2017 Learning surface molecular structures via machine vision npj Comput. Mater. 3 31
[19] Ziatdinov M, Dyck O, Maksov A, Li X, Sang X, Xiao K, Unocic R R, Vasudevan R, Jesse S and Kalinin S V 2017 Deep learning of atomically resolved scanning transmission electron microscopy images: chemical identification and tracking local transformations ACS Nano 11 12742–52
[20] Leitherer A 2023 Byung Christian. Automatic identification of crystal structures and interfaces via artificial-intelligence-based electron microscopy (arXiv:2303.12702)
[21] Akers S, Kautz E, Trevino-Gavito A, Okseta M, Matthews B E, Wang L, Du Y and Spurgeon S R 2021 Rapid and flexible segmentation of electron microscopy data using few-shot machine learning npj Comput. Mater. 7 187
[22] Leitherer A, Ziletti A and Ghiringhelli L M 2021 Robust recognition and exploratory analysis of crystal structures via Bayesian deep learning Nat. Commun. 12 6234
[23] Ziletti A, Kumar D, Scheffer M and Ghiringhelli L M 2018 Insightful classification of crystal structures using deep learning Nat. Commun. 9 2775
[24] Baldi P, Sadowski P and Whiteson D 2014 Searching for exotic particles in high-energy physics with deep learning Nat. Commun. 5 4308
[25] Jean N, Burke M, Xie M, Davis W M, Lobell D B and Ermon S 2016 Combining satellite imagery and machine learning to predict poverty Science 353 790
[26] Litjens G, Sánchez C I, Timofeeva N, Hermens M, Nagtegaal I, Kovacs I, Hulsbergen—van de Kaa C, Bult P, van Ginneken B and van der Laak J 2016 Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis Sci. Rep. 6 26286
[27] Ghosh A, Sumpter B G, Dyck O, Kalinin S V and Ziatdinov M 2021 Ensemble learning and iterative training (ELIT) machine learning: applications towards uncertainty quantification and automated experiment in atom-resolved microscopy (arXiv:2101.08449)
[28] Locatello F, Bauer S, Lucic M, Raetsch G, Söllkopf B and Bachem O 2019 Challenging common assumptions in the unsupervised learning of disentangled representations Int. Conf. on Machine Learning (PMLR) pp 4114–24
[29] Prifti E, Buban J P, Thind A S and Klie R F 2023 Variational convolutional autoencoders for anomaly detection in scanning transmission electron microscopy Small 19 2205977
[30] Zhang R 2019 Making convolutional networks shift-invariant again Int. Conf. on Machine Learning (PMLR) pp 7324–34
[31] Engstrom L, Tsipras D, Schmidt L and Madry A A 2017 Rotation and a translation suffice: fooling CNNs with simple transformations (arXiv:1712.02779)
[32] Goodfellow I, Bengio Y, Courville A and Bengio Y 2016 Deep Learning (MIT Press)
[33] Lin W Z, Li Q, Belianinov A, Sales B C, Sefat A, Gai Z, Baddorf A P, Pan M H, Jesse S and Kalinin S V 2013 Local crystallography analysis for atomically resolved scanning tunneling microscopy images Nanotechnology 24 415707
[34] Belianinov A, He Q, Kravchenko M, Jesse S, Borisevich A and Kalinin S V 2015 Identification of phases, symmetries and defects through local crystallography Nat. Commun. 6 7801
[35] Vasudevan R K, Ziatdinov M, Jesse S and Kalinin S V 2016 Phases and interfaces from real space atomically resolved data: physics-based deep data image analysis Nano Lett. 16 5574–81
[36] Ziatdinov M, Nelson C, Vasudevan R K, Chen D Y and Kalinin S V 2019 Building ferroelectric from the bottom up: the machine learning analysis of the atomic-scale ferroelectric distortions Appl. Phys. Lett. 113 5
[37] Maksov A, Dyck O, Wang K, Xiao K, Geoghegan D B, Sumpter B G, Vasudevan R K, Jesse S, Kalinin S V and Ziatdinov M 2019 Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS2 npj Comput. Mater. 5 8
[38] Ziatdinov M, Dyck O, Li X, Sumpter B G, Jesse S, Vasudevan R K and Kalinin S V 2019 Building and exploring libraries of atomic defects in graphene: scanning transmission electron and scanning tunneling microscopy study Sci. Adv. 5 9
[39] Nord M, Vullum P E, Maclaren I, Tybell T and Holmestad R 2017 Atomap: a new software tool for the automated analysis of atomic resolution images using two-dimensional Gaussian fitting Adv. Struct. Chem. Imaging 3 1
[40] Han J, Go K-J, Jang J, Yang S and Choi Y 2022 Materials property mapping from atomic scale imaging via machine learning based sub-pixel processing npj Comput. Mater. 8 1
[41] Bepler T, Zhong E, Kelley K, Brignole E and Berger B 2019 Explicitly disentangling image content from translation and rotation with spatial-VAE Advances in Neural Information Processing Systems pp 15409–19 (available at: https://proceedings.neurips.cc/paper_files/paper/2019/hash/5a381e824d099699159da10c71c45577-Abstract.html)
[42] Bengio Y, Courville A and Vincent P 2013 Representation learning: a review and new perspectives IEEE Trans. Pattern Anal. Mach. Intell. 35 1798–828
[43] Kingma D P and Welling M 2019 An introduction to variational autoencoders Found. Trends® Mach. Learn. 12 307–92
[44] Kingma D P and Welling M 2013 Auto-encoding variational Bayes (arXiv:1312.6114)
[45] Locatello F, Tschannen M, Bauer S, Rätsch G, Schölkopf B and Bachem O 2019 Disentangling factors of variation using few labels (arXiv:1912.13735)
[46] Landau L D and Lifshitz E M 1976 Mechanics: Volume I (Butterworth-Heinemann)
[47] Higgins I, Amos D, Pfau D, Racaniere S, Matthey L, Rezende D and Lerchner A 2018 Towards a definition of disentangled representations (arXiv:1812.02230)
[48] Ziatdinov M, Fujii S, Kiguchi M, Enoki T, Jesse S and Kalinin S V J N 2016 Data mining graphene: correlative analysis of structure and electronic degrees of freedom in graphenic monolayers with defects Nanotechnology 27 495703
[49] Ziatdinov M, Nelson C T, Zhang X, Vasudevan R K, Eliseev E, Morozovska A N, Takeuchi I and Kalinin S V 2020 Causal analysis of competing atomistic mechanisms in ferroelectric materials from high-resolution scanning transmission electron microscopy data npj Comput. Mater. 6 127
[50] Ziatdinov M and Kalinin S 2021 Robust feature disentanglement in imaging data via joint invariant variational autoencoders: from cards to atoms (arXiv:2104.10180)
[51] Kalinin S V, Oxley M P, Valleti M, Zhang J, Hermann R P, Zheng H, Zhang W, Eres G, Vasudevan R K and Ziatdinov M 2020 Deep Bayesian local crystallography (arXiv:2012.07134)
[52] Kalinin S V, Ziatdinov M, Hinkle J, Jesse S, Ghosh A, Kelley K P, Lupini A R, Sumpter B G and Vasudevan R K 2021 Automated and autonomous experiments in electron and scanning probe microscopy ACS Nano 15 16204–27
[53] Shahriari B, Swersky K, Wang Z, Adams R P and Freitas N D 2016 Taking the human out of the loop: a review of Bayesian optimization Proc. IEEE 104 148–75
[54] Kalinin S V, Valleti M, Vasudevan R K and Ziatdinov M 2020 Exploration of lattice Hamiltonians for functional and structural discovery via Gaussian process-based exploration–exploitation J. Appl. Phys. 128 164304
[55] Ziatdinov M, Wong C Y and Kalinin S V Examples for finding simplicity: unsupervised discovery of features, patterns, and order parameters via shift-invariant variational autoencoders (available at: https://github.com/tommycwong/shift-VAE-atomic/tree/1.0.0)