Robust Spectral Anomaly Detection in EELS Spectral Images via Three Dimensional Convolutional Variational Autoencoders

Seyfal Sultanov, James P. Buban, and Robert F. Klie*

Department of Computer Science, University of Illinois Chicago, Chicago, Illinois 60607, USA and
Department of Physics, University of Illinois Chicago, Chicago, Illinois 60607, USA

(Dated: December 10, 2024)

We introduce a Three-Dimensional Convolutional Variational Autoencoder (3D-CVAE) for automated anomaly detection in Electron Energy Loss Spectroscopy Spectrum Imaging (EELS-SI) data. Our approach leverages the full three-dimensional structure of EELS-SI data to detect subtle spectral anomalies while preserving both spatial and spectral correlations across the datacube. By employing negative log-likelihood loss and training on bulk spectra, the model learns to reconstruct bulk features characteristic of the defect-free material. In exploring methods for anomaly detection, we evaluated both our 3D-CVAE approach and Principal Component Analysis (PCA), testing their performance using Fe L-edge peak shifts designed to simulate material defects. Our results show that 3D-CVAE achieves superior anomaly detection and maintains consistent performance across various shift magnitudes. The method demonstrates clear bimodal separation between normal and anomalous spectra, enabling reliable classification. Further analysis verifies that lower dimensional representations are robust to anomalies in the data. While performance advantages over PCA diminish with decreasing anomaly concentration, our method maintains high reconstruction quality even in challenging, noise-dominated spectral regions. This approach provides a robust framework for unsupervised automated detection of spectral anomalies in EELS-SI data, particularly valuable for analyzing complex material systems.

Introduction—High-resolution transmission electron microscopy has become the predominant approach for characterizing a wide range of materials, ranging from 2D materials [1] to superconductors, [2] semiconductors, [3] and catalysts. [4] A particularly powerful approach to materials characterization is the combination of scanning transmission electron microscopy (STEM) [5] with electron energy-loss spectroscopy (EELS), [6] which can measure the local density of states up to single atomic-column resolutions. [7] This approach is often referred to as EELS spectrum imaging (EELS-SI), [8] and the resulting 3-dimensional data cubes contain a detailed map of elemental composition, electronic structure, and bonding at the atomic scale, which is critical for understanding the fundamental properties of condensed matter systems.

Core-loss EELS, which stems from the transition of highly-localized states, such as the *1s* or *2p* states into unoccupied orbitals above the Fermi-level E_F, often exhibit a detailed fine-structure of a particular edge, for example, the oxygen K-edge or a transition metal L-edge, near the edge onset, which reflects the density of unoccupied states near E_F. [6] Subtle changes in this near-edge fine structure are due to changes in the local crystal structure, changes in orbital or spin ordering, valence state changes or the presence of defects/vacancies. These insights are invaluable for exploring phenomena, such as superconductivity, [9] magnetism, [10] and topological states of matter. [11] Furthermore, atomic-column resolved EELS is particularly impactful in analyzing interfaces, grain boundaries, [12] and low-dimensional materials, [13] where local electronic and chemical environments dictate macroscopic material properties. By bridging the gap between atomic-scale phenomena and bulk material behavior, this technique plays a crucial role in advancing the design of quantum materials, catalysts, and energy devices.

Existing EELS-SI data analysis methods predominantly rely on manual inspection or dimensionality reduction techniques, such as Principal Component Analysis (PCA). While effective at noise reduction and extracting statistically significant features, PCA's linear nature limits its ability to capture physically significant, intricate spectral details. Its variance-based decomposition often relegates subtle spectral features to low-variance components, which are commonly discarded as noise. Furthermore, PCA, being constrained to linear combinations of input features, cannot accurately represent non-linear relationships in the data, potentially overlooking complex spectral patterns crucial for anomaly detection.

Machine learning (ML) has emerged as a significant tool across scientific disciplines, offering new approaches for analyzing complex datasets [14–16]. In electron microscopy, conventional ML techniques have enhanced data analysis, enabling robust methods for denoising images and identifying atoms/patterns in STEM/STM/AFM data [17–19]. The increasing accessibility of high-performance computing has accelerated the adoption of more complex, data-intensive methods, particularly Deep Learning (DL) models like autoencoders, which have gained prominence in physics applications. Autoencoders, hourglass-shaped feed-forward neural networks, compress input data through an encoder, then reconstruct it from a low-dimensional representation, preserving salient features while finding a succinct data representation [20].

Variational Autoencoders (VAEs) [21] combine varia-

tional inference and autoencoders to create deep generative models trainable in an unsupervised fashion. VAEs excel at learning compact, non-linear representations of high-dimensional data. They achieve this by regularizing the latent space so that nearby points encode semantically similar information. This regularization is accomplished by modeling the latent space as a product of Gaussian distributions and minimizing the Kullback-Leibler divergence between the estimated and true underlying distributions [21]. The Kullback-Leibler divergence is minimized when the estimated distribution matches the true underlying distribution, allowing the VAE to learn a smooth, continuous latent space that captures meaningful data features.

Unlike conventional autoencoders, which may learn discontinuous or arbitrary latent representations, VAEs' regularized latent space improves the quality of both learned features and the learned relationships between them. This characteristic makes VAEs more resistant to learning undesirable features such as noise signatures or subtle shifts in the training set - issues that often reduce the semantic meaning of latent encodings in conventional autoencoders - thereby enhancing VAEs' generalizability to new data.

In physics, VAEs have shown ability to learn physically relevant representations. For example, in molecular systems, VAEs have been applied to represent free energy surfaces (FES), enabling improved sampling of high-dimensional spaces and prediction of properties like isothermal compressibility or NMR spin-spin J couplings [22]. They have also been used for dimensionality reduction, such as identifying slowly varying collective variables in peptide folding, which is crucial for developing Markov state models of conformational changes [22].

In materials science, VAEs and related autoencoder architectures have been applied to extract meaningful features from various scientific images, including spatial-spectral characteristics from hyperspectral images using 3D convolutional autoencoders [23] and structural patterns from STEM/STM images using shift-invariant VAEs [24]. The latent space variables often correlate with key physical properties such as atomic positions, lattice periodicities, or electronic states, providing insights into the underlying physics. VAEs have demonstrated the capability to separate individual structural building blocks from relevant order parameter fields that change slowly on the length scale of the atomic lattice, enabling efficient exploration of complex configurational spaces [24].

Recent studies have applied ML to EELS data, creating models for predicting individual spectra from structural images based on the idea that local structures and functional phenomena are correlated through a small number of latent mechanisms [25]. Denoising autoencoders have been explored as an alternative to PCA, matching and outperforming PCA reconstructions [26]. However, most approaches have primarily addressed in-dividual EELS spectra, leaving the full potential of 3D EELS-SI data unexplored.

VAEs have demonstrated effectiveness in anomaly detection across various domains. In civil engineering, they have been applied to detect temporal and spatial anomalies in dam monitoring data [27]. In computer vision, VAEs have been used to detect and localize anomalous events in surveillance videos using only normal samples for training [28]. In medical imaging, 3D VAEs have shown promise in detecting schizophrenia from brain MRI data [29].

Building on these successes, we previously developed an approach using Convolutional VAEs (CVAEs) to detect and classify point defects and other structural anomalies in atomic-resolution STEM images [30, 31]. We successfully validated this method on STEM images of $SrTiO_3$ and more complex structures like $FePO_4$ and $CdTe$. In this approach, a CVAE trained solely on bulk crystal structure images learned the expected atomic positions and intensities. Anomalies were then identified by subtracting the input images from the CVAE's reconstructions.

The present work extends this concept to EELS-SI data, introducing a novel Three-Dimensional Convolutional Variational Autoencoder (3D-CVAE) for discovering intricate spectral anomalies. This unsupervised approach can learn complex, disentangled patterns in EELS-SI data while requiring relatively small training datasets compared to most supervised neural networks. Importantly, it can be implemented using computing resources widely available to researchers in the field, without requiring high performance supercomputers.

To enhance scalability and applicability, our model operates directly on EELS-SI data, eliminating the need for additional feature engineering or comprehensive prior knowledge of the material system. This element-agnostic approach allows the model to learn underlying spectral patterns for any element, given sufficient training examples.

We evaluate our method by training the 3D-CVAE on bulk EELS-SI data and testing its performance in reconstructing spectra with eight different types of artificially injected anomalies. Our results demonstrate that the 3D-CVAE-based method outperforms traditional PCA in both spectral reconstruction and anomaly detection.

For our experiments, we used EELS-SI spectra from epitaxial $BiFeO_3$ thin films grown on $SrTiO_3$

This approach successfully identifies subtle spectral changes associated with defect structures and interface phenomena, surpassing the capabilities of conventional analysis methods.

Methods—The 3D-CVAE employs 3D convolutional layers to simultaneously capture spectral features and their spatial relationships within the EELS-SI datacube. Translational invariance is achieved through strided convolutions [32] across all three dimensions, ensuring con-
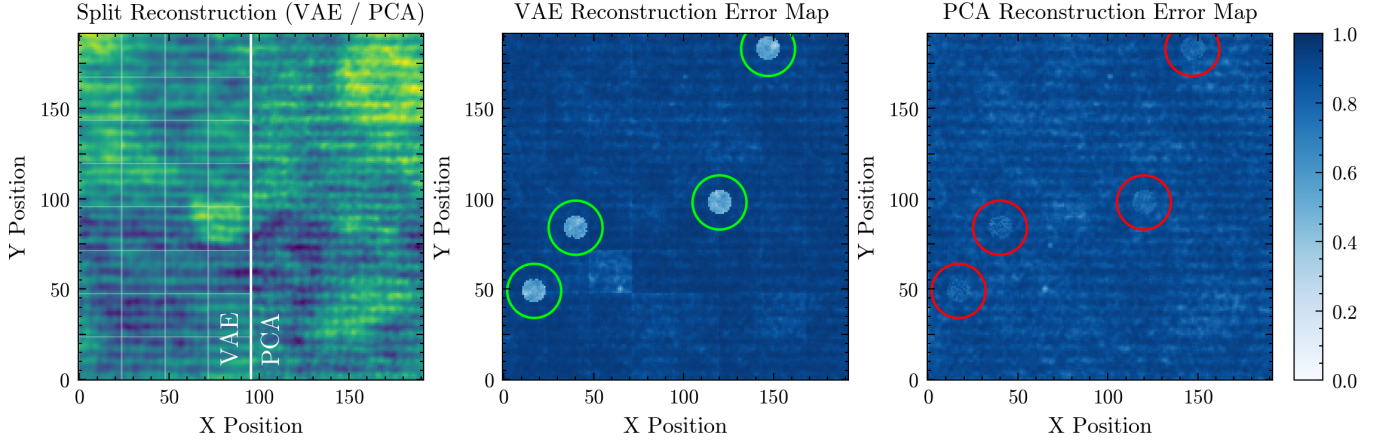
FIG. 1. Comparison of VAE and PCA reconstructions and their anomaly detection performance. (a) Split visualization of the EELS-SI datacube reconstruction, with VAE (left) and PCA (4 components, right) results shown as z-direction intensity sums. (b) VAE reconstruction error heatmap showing Pearson Correlation Coefficients between original and reconstructed spectra. (c) Corresponding PCA reconstruction error heatmap. In both (b) and (c), green circles indicate successfully detected anomalies using Otsu's thresholding method, while red circles mark undetected anomalous regions. Lower PCC values (lighter colors) indicate greater deviation between original and reconstructed spectra.

sistent feature detection regardless of the exact position of spectral features. This architecture efficiently processes the full three-dimensional structure of the data while maintaining spatial relationships.

During training, the 3D-CVAE constructs a low-dimensional latent space by encoding input data as distributions in a continuous latent manifold, regularized by a Kullback-Leibler (KL) divergence term to ensure smoothness and prevent overfitting. The model encodes each spectrum as a point in this latent space, where similar spectra cluster together and variations are captured by the learned manifold structure. During inference, when presented with an anomalous image, the VAE's encoder projects it onto this learned manifold of normality in the latent space [21]. The subsequent reconstruction by the decoder is based on this projection, effectively filtering out features that deviate from the learned normal patterns. This process can be understood as a form of probabilistic dimensionality reduction followed by a generative reconstruction, where the model's learned prior acts as a constraint that guides the reconstruction towards normality. Consequently, the reconstructed image tends to exclude or attenuate anomalous elements, as they are not well-represented in the model's learned distribution of normal features. The discrepancy between the original anomalous input and its "normalized" reconstruction can then serve as a quantitative measure of anomaly, making VAEs an effective tool for both detecting and localizing anomalies in complex, high-dimensional data such as images.

We present a novel DL method for EELS data by reformulating the reconstruction problem through Cross-Entropy Loss (CE). While existing DL approaches to spectral data typically employ Mean Squared Error

(MSE) [26] or Evidence Lower Bound (ELBO) [21] objectives that treat spectral intensities as continuous values, our formulation recognizes the discrete nature of electron energy loss events. Each spectrum in EELS represents a distribution of discrete electron counting events across energy channels. By utilizing Cross-Entropy Loss (CE) instead of MSE or ELBO, we treat each energy channel as a distinct class, where the normalized spectrum intensities represent the probabilities of electron energy loss events. This formulation aligns more closely with the probabilistic nature of the data and improves the model's ability to capture and reconstruct critical spectral features. The total loss function used for training combines the CE Loss term with a Kullback-Leibler (KL) divergence term, as follows:

$$L_{\text{total}}(\mathbf{x}, \hat{\mathbf{x}}) = \underbrace{\sum_{i=1}^{N} L_{\text{CE}}(\mathbf{y}_i, \hat{\mathbf{y}}_i)}_{\text{Cross-Entropy Loss}} + \beta \cdot L_{\text{KL}} \qquad (1)$$

Where $\mathbf{x}$ represents the input shard of the (SI) datacube, and $\hat{\mathbf{x}}$ represents the reconstructed shard produced by the decoder. The total number of spectra in a shard is denoted by N, which is obtained by flattening the spatial dimensions (x, y) of the SI datacube. The parameter β [33] is a weighting factor that controls the trade-off between the reconstruction accuracy (governed by the CE Loss) and the regularization of the latent space (enforced by the KL divergence term). The CE Loss quantifies the discrepancy between the original spectra and their reconstructions. For an individual spectrum, the CE Loss is

defined as:

$$L_{\mathrm{CE}}(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{e=1}^{E} \{ y_e \cdot \log[\mathrm{softmax}(\hat{y}_e)] \} \qquad (2)$$

where $\mathbf{y} = \{y_e\}_{e=1}^{E}$ represents the normalized intensities of the original spectrum and $\hat{\mathbf{y}} = \{\hat{y}_e\}_{e=1}^{E}$ represents the reconstructed normalized intensities. The number of energy channels in each spectrum is denoted by E. The softmax function, $\mathrm{softmax}(\hat{y}_e)$, normalizes the reconstructed intensities to ensure that they are treated as probabilities, with values that sum to 1 across all energy channels. This formulation treats each energy channel as a distinct class, where the original normalized intensities y_e represent the probability of observing an electron energy loss event in channel e. By optimizing this loss, the model reconstructs the spectra in a way that matches the probabilistic distribution of the original input data. In addition to the reconstruction loss, the KL divergence term regularizes the organization of the latent space, ensuring that it is smooth and aligned with a prior Gaussian distribution,. The KL divergence is defined as:

$$L_{\mathrm{KL}} = -\frac{1}{2} \sum_{j=1}^{J} \left[1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2 \right] \qquad (3)$$

Here, J is the dimensionality of the latent space. The terms μ_j and σ_j^2 are the mean and variance of the approximate posterior distribution for the j-th latent dimension, respectively. This term encourages the latent representations to be close to the standard Gaussian prior, promoting a compact and well-organized latent space. The parameter β in the total loss function governs the balance between the strength of this regularization and the fidelity of spectral reconstructions. Higher values of β [33] enforce stricter regularization at the cost of reconstruction accuracy, while lower values prioritize precise reconstructions of the input spectra [21].

Results—To validate our approach, we inject synthetic anomalies in the form of Fe *L*-edge peak shift anomalies in spatially clustered patterns, simulating realistic defect structures. The Fe *L*-edge was specifically chosen for this proof of concept due to its characteristically high Signal-to-Noise Ratio (SNR).

The injected anomalies consist of 2.5 eV peak shifts, chosen to represent realistic defect-induced changes in electronic structure. Figure 2 demonstrates an example of such an anomaly, showing the original Fe L-edge spectrum (black) compared to the anomaly-injected spectrum (red), highlighting the characteristic peak shift our model aims to detect. To evaluate detection performance, we compare our VAE-based approach against PCA reconstructions. The analysis pipeline processes the anomaly-injected datacube through both methods. For the VAE analysis, we segment the datacube into 24×24×L blocks (where L represents the spectral dimension), process these through the network, and recombine

them to preserve the original dimensions. For quantitative comparison, we calculate the Pearson Correlation Coefficient (PCC) between original and reconstructed spectra within the Fe L-edge energy window (690-730 eV) for both methods. The PCC [34] metric measures the linear correlation between two variables, ranging from -1 to 1, and is given by:

$$\rho_{x,y} = \frac{\sum_{e=1}^{E}(x_e - \bar{x})(y_e - \bar{y})}{\sqrt{\sum_{e=1}^{E}(x_e - \bar{x})^2} \cdot \sqrt{\sum_{e=1}^{E}(y_e - \bar{y})^2}} \qquad (4)$$

In this context, the spectrum is treated as a pair of multivariate data vectors, $\mathbf{x} = (x_1, x_2, \ldots, x_E)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_E)$, where E denotes the total number of energy channels. The variables x_e and y_e represent the intensities at the e-th energy channel for the respective spectra. The mean intensities of the spectra are given by $\bar{x} = \frac{1}{E}\sum_{e=1}^{E} x_e$ and $\bar{y} = \frac{1}{E}\sum_{e=1}^{E} y_e$, which capture the average intensity across all energy channels in each spectrum.

Figure 1 provides a comprehensive visualization of both methods' performance highlighting the true positive areas detected by the VAE. To quantitatively identify anomalies, we analyze the distribution of Pearson Correlation Coefficient (PCC) scores across all pixels. As shown in Figure 1, the VAE-generated error maps provide clearer visualization of localized errors, while the corresponding PCC distributions (Figure 3) exhibit distinct bimodality, effectively separating normal and anomalous pixels. In contrast, while PCA-generated error maps show lower mean PCC values for anomalous regions, the distribution lacks clear separation between normal and anomalous populations, making reliable classification challenging. For automated anomaly classification, we implement Otsu's method [35], an algorithm that optimally separates the PCC histogram into two classes by maximizing between-class variance. To minimize false positives in anomaly-free data, we incorporate a unimodality check of the PCC distribution. In the specific case presented in Figures 1 and 3, our VAE approach achieved high classification accuracy, with only
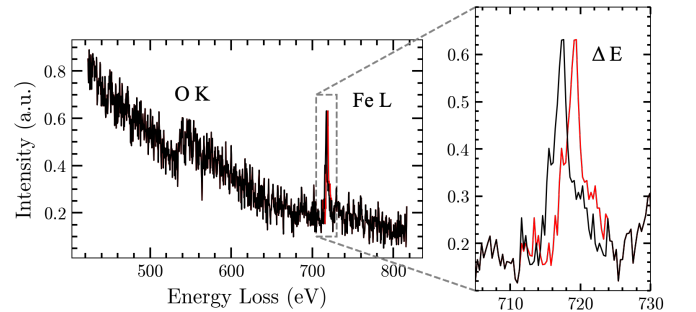

FIG. 2. Example of an injected peak shift anomaly in EELS spectra. The original Fe L-edge spectrum (black) compared to an artificially introduced 2.5 eV peak shift (red segment)

6 anomalous pixels misclassified as normal out of 38,000 total pixels. To verify the robustness of our method, we conducted a statistical analysis comparing our approach with PCA using 3, 4, and 5 components across various shift magnitudes (Figure 4). Performance was evaluated using F1-scores [15, 36], a harmonic mean of precision and recall that balances detection accuracy by accounting for both undetected anomalies (false negatives) and misclassified normal pixels (false positives). Our VAE approach demonstrates consistently high performance, maintaining robust F1-scores across different shift magnitudes with both high precision and recall. PCA performance shows a fundamental trade-off: using three components provides the best anomaly detection among PCA variants but exhibits periodic performance fluctuations based on shift-basis vector alignment. Adding more components improves spectral reconstruction fidelity but degrades anomaly detection capability, a limitation most apparent when examining subtle spectral features such as the O K edge. Further analysis reveals that PCA performance is optimal when anomalies are small in number and sparsely distributed, while our VAE approach maintains consistent performance even beyond physically realistic anomaly concentrations. Although peak shifts be-

FIG. 4. Performance comparison across different magnitudes of peak shifts. F1-scores for VAE (red) and PCA with 3 (blue), 4 (orange), and 5 (green) components. VAE maintains consistently high F1-scores across all shift magnitudes, while PCA exhibits periodic fluctuations in performance. PCA with 3 components shows the best performance among PCA variants, though its effectiveness varies with shift magnitude.

yond 4 eV exceed typical physical scenarios, we extended our analysis to larger shifts to demonstrate that PCA's trend of improving performance from 0 to 4 eV does not persist beyond this threshold and to show periodic nature of PCA performance. To gain insight into the network's internal representations, we analyzed the latent space encodings of 64 pairs of EELS-SI sub-images, where each pair consisted of an unmodified datacube shard and its anomaly-injected counterpart. Analysis of cosine similarity between the 40-dimensional encodings (corresponding to our model's latent space dimensionality) reveals that the encoder consistently places paired images in close proximity within the latent space, as shown by the high correlation values along the diagonal in Figure 5. This proximity is crucial for our anomaly detection approach, as it demonstrates that the encoder recognizes anomalous and normal spectra as fundamentally the same data point, leading to reconstructions that effectively filter out the anomalous features. This behavior confirms that our encoder successfully learns to represent the underlying normal spectral features while being robust to anomalous variations.

Conclusion—We have demonstrated a novel approach for automated anomaly detection in EELS-SI data using a 3D Convolutional Variational Autoencoder. Our method consistently outperforms traditional PCA-based approaches across various shift magnitudes while preserving spectral detail fidelity, though this performance gap narrows with decreasing anomaly concentration. Analysis of the latent space representations reveals that the model develops sophisticated encoding strategies that adapt to local spectral features, enabling robust anomaly detection without compromising reconstruction quality. While manual analysis confirms that our VAE
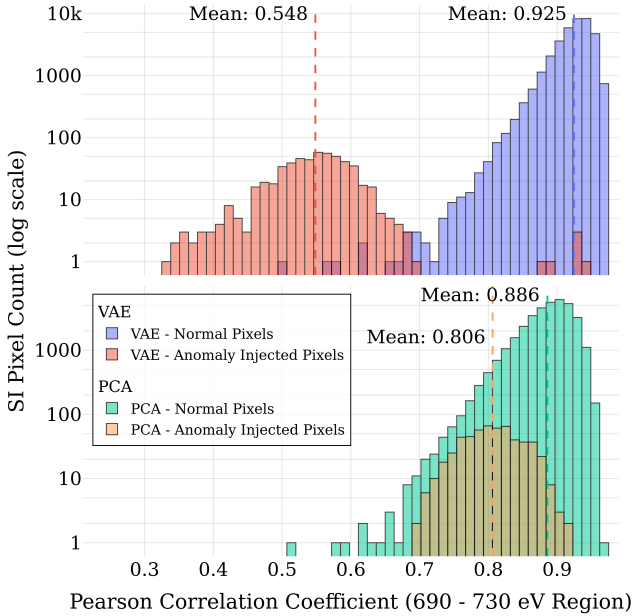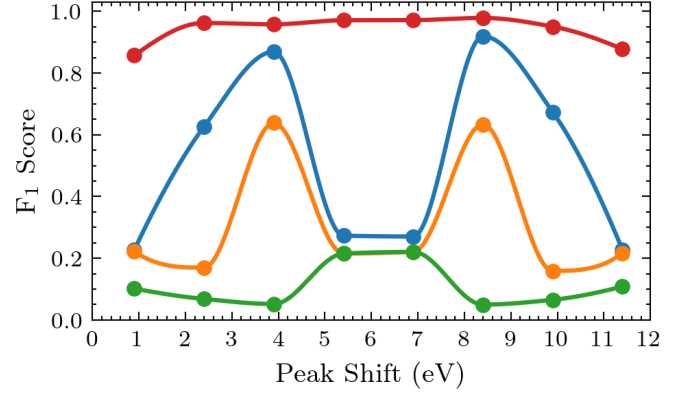

FIG. 3. Distribution of pixels across Pearson Correlation Coefficient (PCC) values for VAE (top) and PCA with 4 components (bottom). Each histogram shows the distribution of normal (bulk) and anomalous pixels on a logarithmic scale. PCC values range from 0.2 to 1.0, where 1.0 indicates perfect correlation between original and reconstructed spectra.The VAE shows clear bimodal separation between normal and anomalous distributions, enabling reliable anomaly detection, while PCA distributions remain overlapped.
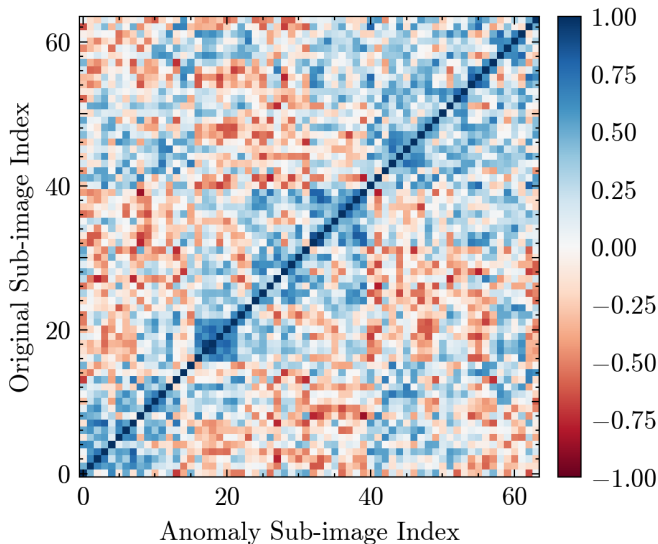
FIG. 5. Visualization of latent space relationships through cosine similarity between 40-dimensional encodings of EELS-SI sub-image pairs. Each point compares an unmodified image (Y axis) with its anomaly-injected counterpart (X axis). The diagonal values close to 1 demonstrate that the encoder places paired images in nearly identical positions in the latent space, confirming that our model successfully recognizes anomalous spectra as variants of their normal counterparts.

approach maintains high reconstruction quality and feature preservation in lower SNR regions such as the O K edge, challenges remain regarding the development of quantitative metrics that can reliably assess reconstruction performance in these noise-dominated spectral regions. Future work will focus on establishing robust evaluation metrics that can better capture the demonstrated capabilities of our method, particularly for subtle spectral features where traditional pixel-by-pixel comparison metrics become unreliable due to noise. Additionally, recent advances in generative AI, particularly stable diffusion [37] models, offer promising directions for EELS denoising and reconstruction. These models' ability to learn complex noise patterns and generate high-quality samples could potentially improve the recovery of fine spectral features in low-SNR regions.

* rfklie@uic.edu

[1] L. Song, L. J. Ci, H. Lu, P. B. Sorokin, C. H. Jin, J. Ni, A. G. Kvashnin, D. G. Kvashnin, J. Lou, B. I. Yakobson, and P. M. Ajayan, Nano Letters **10**, 3209 (2010).

[2] R. F. Klie, J. P. Buban, M. Varela, A. Franceschetti, C. Jooss, Y. Zhu, N. D. Browning, S. T. Pantelides, and S. J. Pennycook, Nature **435**, 475 (2005).

[3] P. M. Voyles, D. A. Muller, J. L. Grazul, P. H. Citrin, and H. J. L. Gossmann, Nature **416**, 826 (2002).

[4] K. Sun, J. Liu, N. Nag, and N. D. Browning, Catalysis Letters **84**, 193 (2002).

[5] S. J. Pennycook and L. A. Boatner, Nature **336**, 565 (1988).

[6] R. Egerton, *Electron Energy Loss Spectroscopy in the Electron Microscope*, 2nd ed. (Springer Science Business Media, New York, 2011).

[7] M. Varela, S. D. Findlay, A. R. Lupini, H. M. Christen, A. Y. Borisevich, N. Dellby, O. L. Krivanek, P. D. Nellist, M. P. Oxley, L. J. Allen, and S. J. Pennycook, Physical Review Letters **92**, 095502 (2004).

[8] C. Jeanguillaume and C. Colliex, Ultramicroscopy **28**, 252 (1989).

[9] N. D. Browning, M. F. Chisholm, S. J. Pennycook, D. P. Norton, and D. H. Lowndes, Physica C **212**, 185 (1993).

[10] R. F. Klie, J. C. Zheng, Y. Zhu, M. Varela, J. Wu, and C. Leighton, Physical Review Letters **99**, 047203 (2007).

[11] M. D. Li, C. Z. Chang, L. J. Wu, J. Tao, W. W. Zhao, M. H. W. Chan, J. S. Moodera, J. Li, and Y. M. Zhu, Physical Review Letters **114** (2015), 10.1103/PhysRevLett.114.146802.

[12] R. F. Klie and N. D. Browning, Applied Physics Letters **77**, 3737 (2000).

[13] F. Lagunas, D. Bugallo, F. Karimi, Y. J. Yang, H. O. Badr, J. H. Cope, E. Ferral, M. W. Barsoum, Y. J. Hu, and R. F. Klie, Chemistry of Materials **36**, 2743 (2024).

[14] M. H. Mobarak, M. A. Mimona, M. A. Islam, N. Hossain, F. T. Zohura, I. Imtiaz, and M. I. H. Rimon, Applied Surface Science Advances **18**, 100523 (2023).

[15] Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).

[16] M. I. Jordan and T. M. Mitchell, Science **349**, 255 (2015).

[17] R. Lin, R. Zhang, C. Wang, X. Q. Yang, and H. L. Xin, Scientific Reports **11** (2021), 10.1038/s41598-021-84499-w.

[18] Y. Hui and Y. Liu, (2018).

[19] S. Somnath, C. R. Smith, S. V. Kalinin, M. Chi, A. Borisevich, N. Cross, G. Duscher, and S. Jesse, Advanced Structural and Chemical Imaging **4** (2018), 10.1186/s40679-018-0052-y.

[20] G. E. Hinton and R. R. Salakhutdinov, Science **313**, 504 (2006).

[21] D. P. Kingma and M. Welling, (2013), original VAE paper.

[22] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Reviews of Modern Physics **91**, 045002 (2019).

[23] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, IEEE Transactions on Geoscience and Remote Sensing **57**, 6808 (2019).

[24] M. Ziatdinov, C. Y. T. Wong, and S. V. Kalinin, Machine Learning: Science and Technology **4** (2023),

10.1088/2632-2153/ad073b.

[25] M. Ziatdinov, A. Ghosh, T. Wong, and S. V. Kalinin, Nature Machine Intelligence **4**, 1101 (2022).

[26] C. M. Pate, J. L. Hart, and M. L. Taheri, Scientific Reports **11**, 19515 (2021).

[27] X. Shu, T. Bao, Y. Zhou, R. Xu, Y. Li, and K. Zhang, Structural Health Monitoring **22**, 39 (2023).

[28] Y. Fan, G. Wen, D. Li, S. Qiu, M. D. Levine, and F. Xiao, Computer Vision and Image Understanding **195**, 102920 (2020).

[29] H. Yamaguchi, Y. Hashimoto, G. Sugihara, J. Miyata, T. Murai, H. Takahashi, M. Honda, A. Hishimoto, and Y. Yamashita, Frontiers in Neuroscience **15** (2021), 10.3389/fnins.2021.652987.

[30] E. Prifti, J. P. Buban, A. S. Thind, and R. F. Klie, Small **19** (2023), 10.1002/smll.202205977.

[31] R. A. W. Ayyubi, J. P. Buban, and R. F. Klie, Microscopy and Microanalysis **30** (2024), 10.1093/mam/ozae044.180.

[32] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Proceedings of the IEEE **86**, 2278 (1998).

[33] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, in *International Conference on Learning Representations* (2017).

[34] K. Pearson and F. Galton, Proceedings of the Royal Society of London **58**, 240 (1895).

[35] N. Otsu, IEEE Transactions on Systems, Man, and Cybernetics **9**, 62 (1979).

[36] C. Van Rijsbergen, in *Proceedings of the joint IBM/University of Newcastle upon tyne seminar on data base systems*, Vol. 79 (1979) pp. 1–14.

[37] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," (2022), arXiv:2112.10752 [cs.CV].

[38] A. Biswas, M. Ziatdinov, and S. V. Kalinin, Machine Learning: Science and Technology **4**, 045004 (2023).

[39] J. Sun, X. Wang, N. Xiong, and J. Shao, IEEE Access **6**, 33353 (2018).

[40] (2019).

[41] S. J. Pennycook and P. D. Nellist, eds., *Scanning Transmission Electron Microscopy* (Springer New York, 2011).

[42] J. An and S. Cho, "Snu data mining center 2015-2 special lecture on ie variational autoencoder based anomaly detection using reconstruction probability," (2015), in the notebook.

[43] T. Matsuo, H. Fukuhara, and N. Shimada, (2017).

[44] K.-K. Ng and M.-F. Yang, Physical Review B **108**, 214428 (2023).

[45] Z. Cheng, E. Zhu, S. Wang, P. Zhang, and W. Li, IEEE Access **9**, 43991 (2021).

[46] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," (2015).
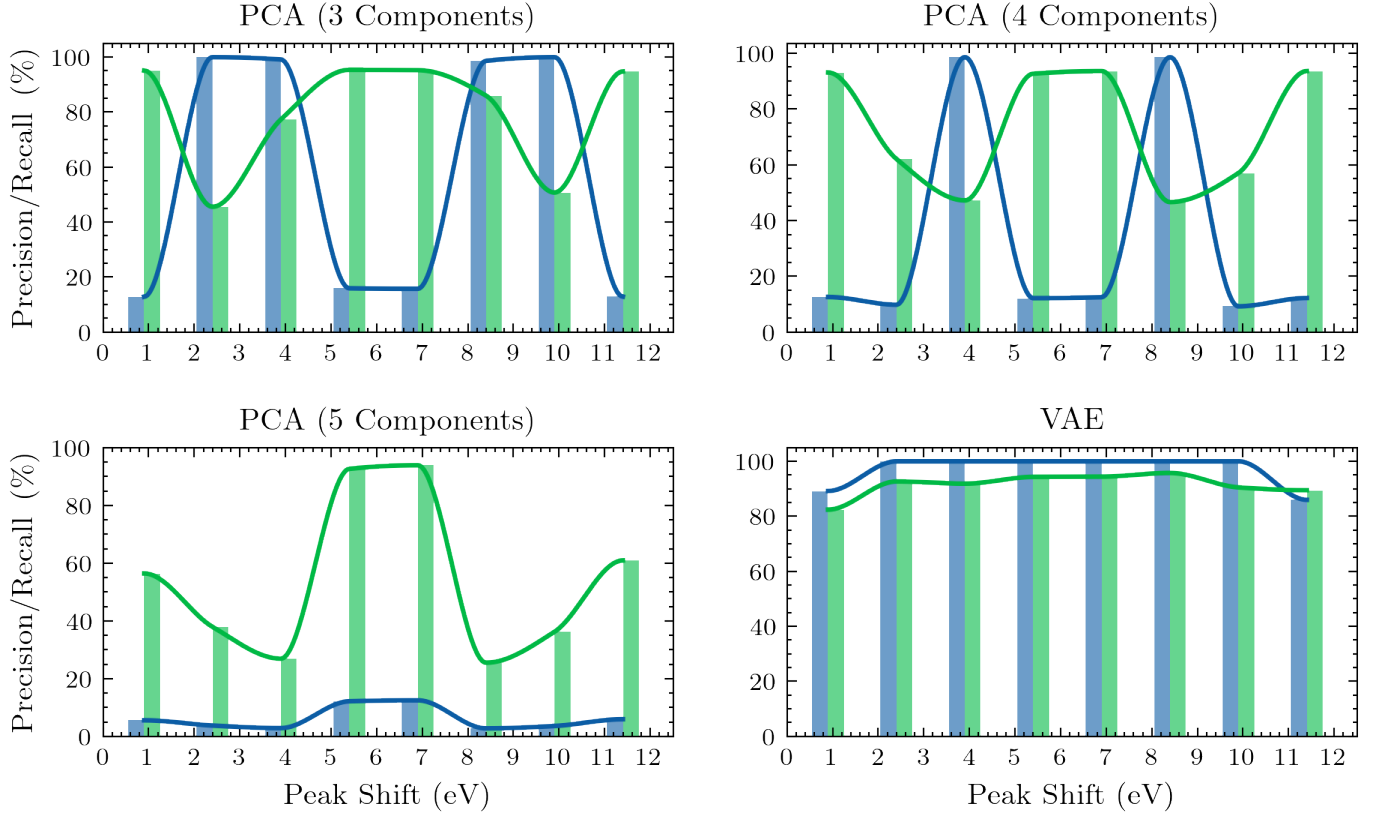
END MATTER

FIG. 6. Precision-recall curves comparing anomaly detection performance between our 3D-CVAE model and PCA with 3,4,5 components. The VAE demonstrates consistently superior performance across different detection thresholds, maintaining a better balance between precision and recall. In contrast, PCA exhibits trade-off behavior, where improved precision comes at the cost of substantially reduced recall and vice versa, indicating less reliable anomaly detection.
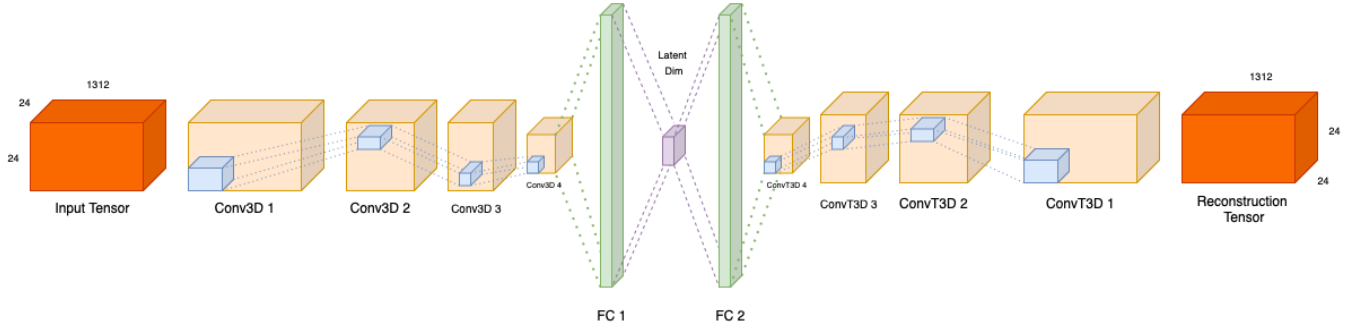

FIG. 7. Schematic representation of the 3D-CVAE architecture. The encoder processes the input EELS SI datacube through convolutional layers to a compressed latent space representation. The decoder then reconstructs the data from this latent space back to the original dimensionality.