Machine Learning Data Augmentation Strategy for Electron Energy Loss Spectroscopy: Generative Adversarial Networks

Daniel del-Pozo-Bueno, Demie Kepaptsoglou, Quentin M Ramasse, Francesca Peiró, Sònia Estradé

Microscopy and Microanalysis, 2024, **30**, 278–293
https://doi.org/10.1093/mam/ozae014

Original Article

Microscopy AND Microanalysis

Machine Learning Data Augmentation Strategy for Electron Energy Loss Spectroscopy: Generative Adversarial Networks

Daniel del-Pozo-Bueno[1,2,*], Demie Kepaptsoglou[3,4], Quentin M. Ramasse[3,5], Francesca Peiró[1,2], and Sònia Estradé[1,2]

[1]*LENS-MIND, Departament d'Enginyeria Electrònica i Biomèdica, Universitat de Barcelona, 1-11 Martí i Franquès, 08028 Barcelona, Spain*
[2]*Institute of Nanoscience and Nanotechnology (IN2UB), Universitat de Barcelona, 1-11 Martí i Franquès, 08028 Barcelona, Spain*
[3]*SuperSTEM Laboratory, Sci-Tech Daresbury Campus, Keckwick Lane, Daresbury WA4 4AD, UK*
[4]*School of Physics, Engineering and Technology, University of York, Newton way, YO10 5DD Heslington, UK*
[5]*Schools of Chemical and Process Engineering & Physics and Astronomy, Woodhouse Lane, University of Leeds, LS2 9JT Leeds, UK*

***Corresponding author:** Daniel del-Pozo-Bueno, E-mail: dpb.523@gmail.com

Abstract

Recent advances in machine learning (ML) have highlighted a novel challenge concerning the quality and quantity of data required to effectively train algorithms in supervised ML procedures. This article introduces a data augmentation (DA) strategy for electron energy loss spectroscopy (EELS) data, employing generative adversarial networks (GANs). We present an innovative approach, called the data augmentation generative adversarial network (DAG), which facilitates data generation from a very limited number of spectra, around 100. Throughout this study, we explore the optimal configuration for GANs to produce realistic spectra. Notably, our DAG generates realistic spectra, and the spectra produced by the generator are successfully used in real-world applications to train classifiers based on artificial neural networks (ANNs) and support vector machines (SVMs) that have been successful in classifying experimental EEL spectra.

Key words: data augmentation, electron energy loss spectroscopy, generative adversarial networks, machine learning, support vector machines

Introduction

Since the advent of spherical aberration correctors in the early 2000s, there have been substantial improvements in both spatial and energy resolutions of scanning and transmission electron microscopy (S/TEM) and electron energy loss spectroscopy (EELS) techniques because of several technological advancements, such as monochromators, direct detectors, and increased computational power. However, another challenge has arisen from these instrumental enhancements: the generation of vast quantities of data per acquisition. As a result, there is a growing recognition within the EELS community of the necessity for new data analysis tools, such as machine learning (ML) algorithms, to accurately and efficiently process this abundance of data.

When it comes to the confluence of ML and EELS, the current state of the art reveals that the applied strategies can be broadly categorized into two main approaches: supervised and unsupervised. Supervised approaches involve training ML models using labeled data, where the desired output or target is known. These models learn from the labeled examples to make predictions or classify new, unseen data points. In the context of EELS, supervised ML algorithms can be employed for specific tasks such as material identification and elemental mapping by leveraging prelabeled training data. Supervised strategies based on support vector machines (SVMs) and artificial neural networks (ANNs) have been successfully applied to identify the oxidation state of transition metals through their "white line" spectroscopic signatures (Chatzidakis & Botton, 2019; del-Pozo-Bueno et al., 2023).

On the other hand, unsupervised approaches aim to identify patterns, structures, or relationships within the data without the use of explicit labels. These algorithms explore the inherent structure of the EELS data and uncover hidden insights or clusters. Unsupervised ML techniques, such as clustering algorithms or dimensionality reduction methods, can assist in exploratory data analysis and identifying novel features or spectra in EELS data sets. A plethora of studies employed dimensionality reduction algorithms, such as independent component analysis (ICA), principal component analysis (PCA), and negative matrix factorization (NMF), to scrutinize EELS signals (Bonnet et al., 1999; Bonnet & Nuzillard, 2005; de la Peña et al., 2011; Yedra et al., 2012; Torruella et al., 2016; Pelaez-Fernandez et al., 2022). More recent approaches make use of clustering algorithms such as *K*-means, hierarchical clustering (Torruella et al., 2016, 2018), or hierarchical density-based spatial clustering of applications with noise (HDBSCAN) in conjunction with uniform manifold approximation and projection (UMAP) to identify signals in spectral images (Blanco-Portals et al., 2022). Recently, autoencoders

Received: July 29, 2023. **Revised:** August 1, 2023. **Accepted:** February 12, 2024

have also been employed as a strategy for noise removal during the rapid acquisition of EELS spectra (Pate et al., 2021).

Recent advancements in ML have highlighted the crucial role of data quality and preprocessing in achieving accurate models. Whether training supervised algorithms or applying preprocessing in unsupervised algorithms (Maharana et al., 2022), the data used play a vital role in model performance. The requirement for large and diverse training data sets is essential for the development of accurate and general models (Halevy et al., 2009). However, obtaining such data can be costly. Additionally, in supervised models, data labeling prior to model training is needed. Consequently, ensuring both the quality and quantity of data becomes a challenging task when applying ML models.

In this context, data augmentation (DA) by an EELS spectrum generator would prove to be highly advantageous for the EELS community, as it could serve as a key element for the successful implementation of supervised ML strategies in EELS. Primarily, this approach can overcome challenges in data collection for electron beam-sensitive samples. Secondly, generating synthetic spectra simplifies training supervised ML models. Once enough and diverse spectra are acquired to train the generator, it can then produce as much synthetic data as required.

Artificial data generation can be immensely beneficial, especially in situations involving imbalanced data sets (Tanaka & Aranha, 2019), where DA often beats the manual creation of synthetic examples and facilitates the generation of more labeled examples. DA has been already used in other scientific communities, for instance in biomedicine, to generate synthetic noncontrast computed tomography (CT) images from contrast CT images to augment the training data for a U-Net segmentation model (Sandfort et al., 2019); to generate medical images improving the performance of convolutional neural networks (CNNs) for medical image classification (Frid-Adar et al., 2018); or introducing GAN-derived synthetic data into the training data sets for brain segmentation tasks (Bowles et al., 2018).

In this work, we propose an unsupervised strategy based in generative adversarial networks (GANs) aimed at augmenting electron energy loss (EEL) spectra through the generation of synthetic spectra derived from well-known data. Specifically, we develop data augmentation GANs (DAGs) to design a successful architecture capable of generating synthetic EEL spectra from a modest number of experimental spectra—approximately 125 spectra. This approach enables us to create an abundant quantity of synthetic spectra as required to train supervised or semisupervised ML algorithms aimed at classifying real data. We demonstrate the efficacy of our strategy by generating synthetic EEL signals of the transition metals $L_{2,3}$ (white lines) and the oxygen K-edges of iron (Fe) and manganese (Mn)-based oxides. The synthetic spectra, generated through DAGs, are subsequently used to train classifiers to discriminate the oxidation state of Fe and Mn in these metal oxides through their respective white lines.

Materials and Methods

Experimental Data Set Construction

In order to train GAN models, it is necessary to supply the GANs with EEL spectra containing the features of interest for generation. In this regard, we concentrate on the following EELS features of compounds of interest: the transition metal $L_{2,3}$ "white lines" and the oxygen K-edge. In Figure 1a, these edges are visualized for two Fe oxides, wüstite (FeO) and magnetite (Fe_3O_4). The electron energy loss near-edge structure (ELNES, or "fine structure") of the oxygen K-edges can be primarily divided into four distinct spectral features, consisting of a preedge, a main peak, and two subsequent peaks of lower intensity. The primary difference between the two oxidation states is noticeable in the intensity of the prepeak and a slight energy shift in the main peak (Colliex et al., 1991). The Fe white lines are characterized by two edges: the first and most intense corresponding to the L_3 edge, with an onset value of approximately 708 eV, and a less intense second edge known as L_2, with an onset value of around 721 eV. It is important to note that both onset values and L_3–L_2 intensity ratio depend on the oxidation state and the specific compound under study (Colliex et al., 1991; Tan et al., 2012). Similarly, in Figure 1b, the Mn white lines are presented. The first and most intense corresponds to the L_3 edge, with an onset value of approximately 640 eV, and a less intense second edge known as L_2, with an onset value of around 653 eV. In this case, these features are also influenced by their oxidation state, with Mn^{2+} exhibiting the lowest energy loss and Mn^{4+} displaying the highest energy loss (Tan et al., 2012).

Original experimental spectra were obtained from spectrum images (SIs) of magnetite nanocubes (NCs) and wüstite–magnetite core–shell NCs, with only the signal from the nucleus corresponding to wüstite being used. Given that the NCs exhibit a core–shell structure, the wüstite signal obtained is influenced to some extent by the top and bottom shell layers. These NCs were previously studied in the works of Torruella et al. (2016), Muro-Cruces et al. (2019), and del-Pozo-Bueno et al. (2021). The extraction and labeling of these spectra were carried out using the UMAP and HDBSCAN algorithms to ensure the correct labeling of each spectrum (Torruella et al., 2018; Blanco-Portals et al., 2022). The Mn white line spectra were obtained from pure Mn oxides from the work of Chatzidakis & Botton (2019). Additional Fe and Mn white line spectra were obtained from $Fe_3O_4/Mn_xFe_{3-x}O_4$ core–shell NPs, consisting of a magnetite oxide core surrounded by a Mn ferrite shell (Oberdick et al., 2018). As a result, each class within the data set corresponds to either the Mn ferrite shell or the magnetite core. Similar to the previous NCs, these classes were labeled using UMAP and HDBSCAN algorithms.

A total of four experimental data sets were built using the aforementioned spectra (see first and second columns in Table 1). The first data set, WL data set, contained a total of 7,187 spectra composed of 2,744 spectra of wüstite (FeO), 1,350 spectra of magnetite (Fe_3O_4), 992 spectra of MnO, 841 spectra of Mn_2O_3, and 1,260 spectra of MnO_2. The Fe oxide spectra span 700 energy channels (ch) from 612.5 to 787.5 eV with an energy dispersion of 0.25 eV/ch, while the Mn oxide spectra span 700 energy channels from 615 to 685 eV with an energy dispersion of 0.1 eV/ch. All the energy resolutions were measured from the full width at half maximum (FWHM) of the zero loss peak (ZLP), after setting the energy dispersion in the spectrometer.

The second data set, W data set, was composed of 2,744 spectra of wüstite and 1,350 spectra of magnetite. It contained the oxygen K-edge and the Fe white lines for an energy range of 460 to 800 eV, corresponding to 1,360 energy channels with an energy dispersion of 0.25 eV/ch.

Subsequently, the third data set, K data set, was derived from the W data set and contained the spectra with only the oxygen K-edge. This data set was obtained by cropping the

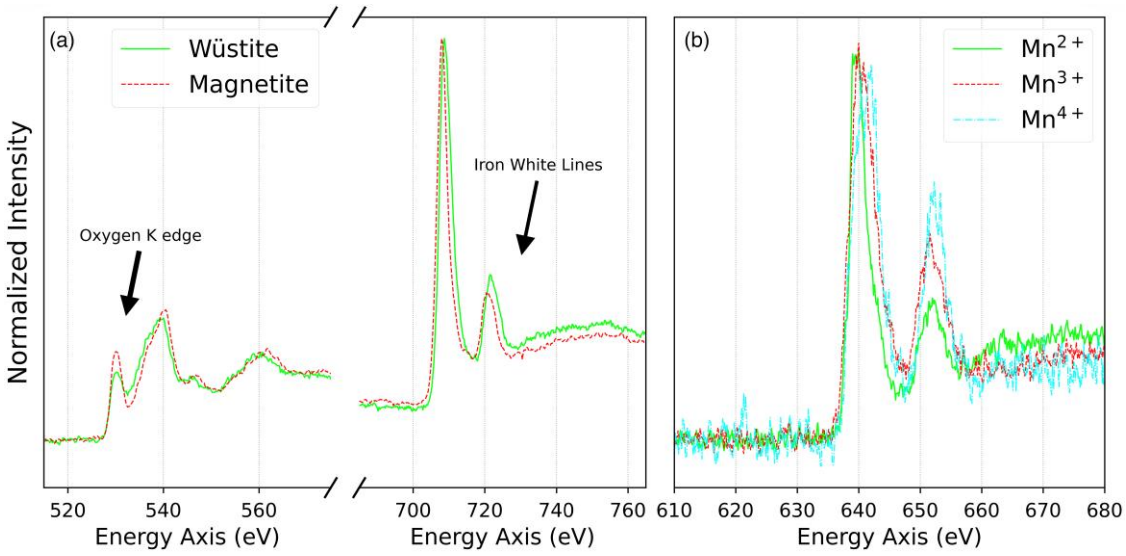


Fig. 1. (**a**) EELS Fe oxide spectra between 515 and 765 eV, containing the oxygen K-edge and Fe white lines. (**b**) EELS spectra of Mn oxides between 610 and 680 eV, containing Mn white lines.

energy range of the W data set from 460 to 610 eV, corresponding to 600 energy channels at the same energy dispersion.

The final data set, known as the MF data set, consisted of 2,458 spectra containing both Fe and Mn white lines and was obtained from Fe–Mn oxide core–shell NPs. It was composed of 1,415 spectra of the magnetite core and 1,043 spectra of the Mn–ferric shell. This data set covers the Mn and Fe white lines within an energy range of 580–784.7 eV, corresponding to 2,048 energy channels, each with an energy dispersion of 0.1 eV/ch.

For all the EEL spectra in these data sets, a straightforward preprocessing procedure was employed. This process involved removing the background prior to the edges of interest and normalizing the resulting spectra using the maximum norm. The background removal was accomplished by the standard methodology in the EELS community using a power-law fitting on the preedge area. It is worth noting that the background removal for the W data set was applied in the oxygen K-edge preedge, while for the MF data set, it was applied from 590 to 620 eV (prior to the Mn white lines). In addition, the MF data set underwent PCA for noise reduction. All of these preprocessing steps were performed using the HyperSpy and Scikit-learn Python modules (de la Peña et al., 2022).

These experimental data sets were then used to feed the DAG models as shown in Figures 2a and 3 (dashed green box). Just 15% of the spectra were used for training, and the rest were used to test the classifiers trained with augmented data.

Data Augmentation Generative Adversarial Network Training Framework

In this study, we employ GAN, which is a class of game theory-based approaches for learning generative models (Goodfellow et al., 2014). This model consists of two networks: a generator network, denoted by G, and a discriminator network, denoted by D, against which the generative model is pitted. This last network learns to distinguish between samples from the actual data distribution and those created by the generative network, as sketched in Figure 2a. The rivalry between the generative and discriminator models self-motivates both to improve their performance until the artificial spectra cannot be distinguished from the actual spectra.

Figure 2b illustrates, within our context of spectra generation, the standard and simplified training scheme employed for a GAN. This involves a sequence of iterative steps that include two concurrent stochastic gradient descents conducted by the Adam optimizer (Kingma & Ba, 2014).

In this particular work, four DAG methodologies are employed for the training of the GAN models as illustrated in Figure 3. These strategies, proposed by Tran et al. (2021), allow for a smaller subset of the data set to be used for training the GAN, thus enabling the production of a functional generator. Under this DAG approach, transformations are applied to both real [$T_i(x)$] and generated [$T_i(G(z))$] spectra before they are submitted to the discriminator, and the number of discriminators used is also varied. Therefore, as depicted in Figure 3, the discriminator or discriminators can receive transformed real and synthetic data or original and generated data. These transformations are designed to enhance the available training data and must be compatible with the data under study. In this context, random energy shifts and random Poissonian noise are added as transformations, denoted as T_1 and T_2, respectively. These transformations are performed alongside the traditional path, which supplies the discriminator with untransformed data. For all the designs, the data without transformation train the discriminator twice for each generator training iteration. The first strategy, known as Single-MonoTrans GAN (Fig. 3a), employs a single discriminator. The second strategy, referred to as Multi-MonoTrans GAN (Fig. 3b), uses two discriminators. The third strategy, Single-BiTrans GAN (Fig. 3c), also employs a single discriminator, but with two transformations. Lastly, the fourth strategy, named Multi-BiTrans GAN (Fig. 3d), utilizes three discriminators and two transformations.

Typically, in GAN training, the models are trained over a large number of epochs until they converge. However, in our case, we wanted to avoid this resource-intensive process. To

Table 1. "Classifier Accuracy by Spectral Feature and DAG Design." This table presents the test accuracy of ANN and SVM classifiers across various datasets (W, WL, K, and MF), utilizing different DAG designs and architectures. Each cell displays the test accuracy percentage corresponding to each combination of dataset, DAG design, and architecture. The accuracy values for ANN and SVM are placed side by side, color-coded with green indicating the best performance and red signifying the worst.

			ANN	SVM
Dataset	DAG Design	Architecture	Test accuracy	
W	Single-MonoTrans	SimpleDense	98,1%	83,2%
		LargeDense	97,3%	85,0%
		SimpleConvo	85,8%	80,7%
	Multi-MonoTrans	SimpleDense	96,8%	94,1%
		LargeDense	50,0%	71,6%
		SimpleConvo	98,9%	97,6%
	Single-BiTrans	SimpleDense	98,8%	92,6%
		SimpleConvo	90,0%	97,7%
	Multi-BiTrans	SimpleDense	99,0%	88,7%
		SimpleConvo	90,8%	96,1%
WL	Single-MonoTrans	SimpleDense	95,0%	93,9%
		LargeDense	89,3%	91,9%
		SimpleConvo	92,6%	83,5%
	Multi-MonoTrans	SimpleDense	91,9%	96,1%
		LargeDense	85,9%	89,0%
		SimpleConvo	95,1%	90,3%
	Single-BiTrans	SimpleDense	96,0%	95,9%
		SimpleConvo	96,4%	90,3%
	Multi-BiTrans	SimpleDense	92,1%	93,3%
		SimpleConvo	87,8%	92,8%
K	Single-MonoTrans	SimpleDense	99,2%	95,6%
		LargeDense	83,0%	80,8%
		SimpleConvo	97,7%	95,3%
	Multi-MonoTrans	SimpleDense	99,2%	96,0%
		LargeDense	96,9%	94,2%
		SimpleConvo	93,7%	92,9%
	Single-BiTrans	SimpleDense	99,1%	92,4%
		SimpleConvo	98,5%	80,9%
	Multi-BiTrans	SimpleDense	99,4%	97,0%
		SimpleConvo	97,3%	96,2%
MF	Single-MonoTrans	SimpleDense	96,3%	98,8%
		LargeDense	94,0%	95,8%
		SimpleConvo	93,2%	89,8%
	Multi-MonoTrans	SimpleDense	93,8%	98,7%
		LargeDense	96,9%	97,6%
		SimpleConvo	96,9%	86,9%
	Single-BiTrans	SimpleDense	98,3%	98,8%
		SimpleConvo	91,4%	97,7%
	Multi-BiTrans	SimpleDense	98,0%	98,0%
		SimpleConvo	90,8%	96,1%

ensure that, we introduced a stopping criterion, which maintained the integrity of the generated data and was computationally efficient, defined by the evaluation of the generated data at each epoch using three distinct metrics: the Fréchet inception distance (FID), the Pearson correlation coefficient (PCC), and the cosine distance (CosD). Similar early stopping strategies have been recently applied in GAN training. For instance, Ji et al. (2023) employed hypothesis testing to assess the alignment between real and generated data distributions, while Ishak et al. (2023) emphasized the importance of validation metrics over traditional convergence approaches with a validation-based method. Regarding these chosen metrics, the FID is a widely used metric for evaluating the performance of generative models (Arora et al., 2017), such as GANs and variational autoencoders, while the other two metrics, PCC and CosD, were chosen for their relevance in evaluating EELS data. For a detailed explanation regarding the evaluation metrics, the reader is referred to the Supplementary Material. The values used to determine when to halt each GAN training session were carefully selected based on a baseline calculated using the distribution of real spectra (Supplementary Table S1).

During GAN training, mode collapse often is a significant challenge, analogous to the overfitting observed in supervised methods. Mode collapse occurs when the generator produces identical or near-identical outputs from different input noise vectors. In this situation, the discriminator's gradient tends to point in the same direction for similar spectra. Given that the discriminator processes each spectrum independently and its gradients are uncoordinated, it fails to guide the generator toward producing diverse spectra. Consequently, all generated spectra tend to resemble a particular spectrum, which the discriminator assumes exceptionally realistic. Once the mode collapse has taken place, the discriminator recognizes this repetitive pattern as the generator's output. However, due to the gradient descents' limitations, it is unable to differentiate between highly similar spectra. As a result, the discriminator's gradients fail to accurately represent this spectrum, hindering the model's convergence toward a spectrum distribution with adequate entropy, i.e., a satisfactory variety of spectra.

To mitigate this issue, we implemented several strategies. First, we employed the two time-scale update rule (TTUR), setting different learning rates for the discriminator and the generator (Heusel et al., 2017). As a result, we trained the discriminator twice per generator iteration with a larger learning rate. Additionally, our training incorporated a one-side label smoothing strategy, which has been shown to enhance the performance of GAN models. Originally proposed by Szegedy et al. (2015), it softens binary targets like 0 and 1 to values like 0.1 and 0.9. When applied one sidedly, as noted by Salimans et al. (2016), it optimally smooths the true value to 0.9, preventing discriminator overconfidence. Finally, another solution to address this collapse mode is to allow the discriminator to examine multiple data instances collectively, which can be accomplished via the technique known as minibatch discrimination (Salimans et al., 2016) that facilitates the discriminator in recognizing a broader spectrum of generated data, thereby mitigating this issue.

The training methodology of the DAG, incorporating the aforementioned strategies, is summarized in the Supplementary Material.

Model Architectures and Convergence: Generator and Discriminator

The conventional approach to GANs typically employs convolutional networks. However, considering the 1D nature of the data, we initially opted for dense networks, as shown in Figures 4 and 5 (Supplementary Figures S1 and S2). To further diversify our approach, we integrated 1D convolutional networks, termed SimpleConvo architecture, ensuring that their design harmonized with the dense networks. These innovative designs for the generator and discriminator were aimed to produce realistic spectra. Alongside, we introduced deeper architectures that retain the foundational structure but incorporate more hidden layers.

More details about the architectures implemented can be found in the Supplementary Material. Nevertheless, here it is worth to mention some of the challenges faced during the construction and training of our DAG models. As already mentioned, one of the primary issues was mode collapse, where

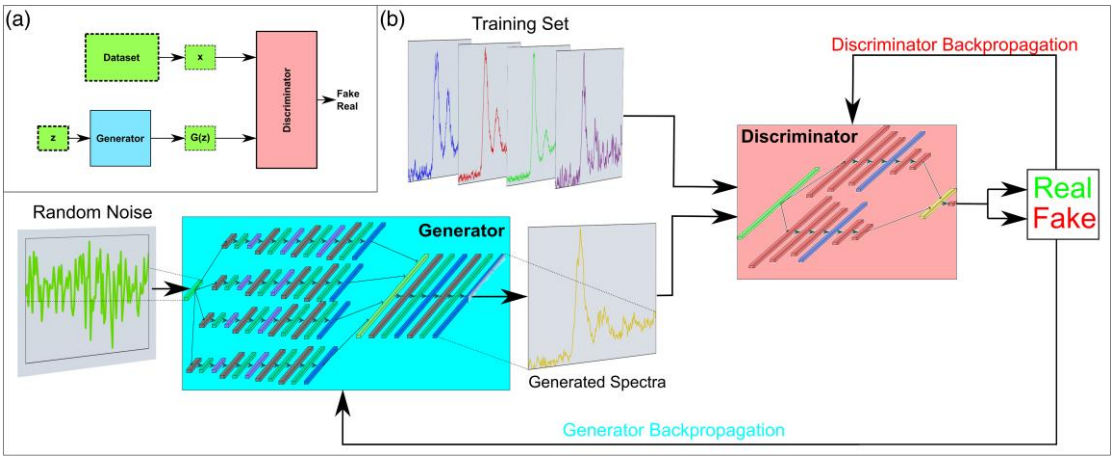


Fig. 2. Illustrations of the vanilla configuration of a GAN: (**a**) generic schematic structure of the GAN strategy and (**b**) GAN configuration in context with the spectra generation task.

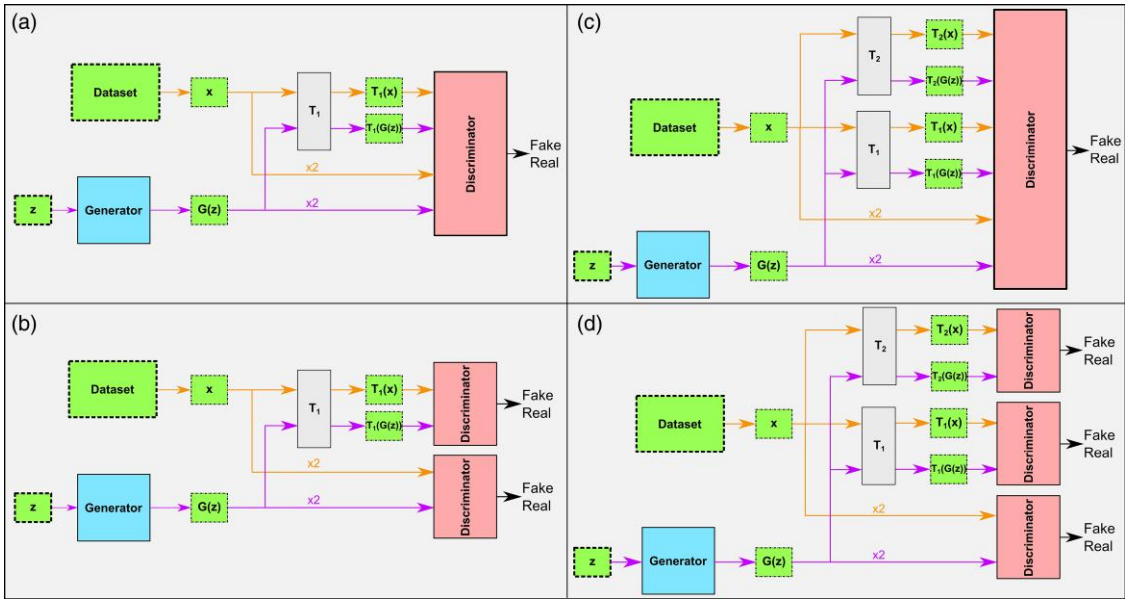


Fig. 3. DAG designs proposed in this work: (**a**) Single-MonoTrans GAN design with a single discriminator and T_1 transformation (adding random energy shifts); (**b**) Multi-MonoTrans GAN design featuring two discriminators and T_1 transformation; (**c**) Single-BiTrans GAN design incorporating a single discriminator along with T_1 and T_2 transformations (adding random Poissonian noise); and (**d**) Multi-BiTrans GAN design utilizing three discriminators with T_1 and T_2 transformations.

our generators were converging to a single point, resulting in virtually all generated spectra being identical with minimal variance. Despite trialing numerous variations of the architecture for both the discriminator and the generator, none was able to prevent the generator from collapsing. The only effective strategy to overcome this issue was the implementation of minibatch discrimination as two intermediate layers in each branch of the discriminator for all the architectures studied (Salimans et al., 2016), as illustrated in Figure 4.

To generate noise distributions with sufficient variation and achieve desirable variance in the generated edges, we integrated two additional layers into the proposed architectures, not included in the Keras module. These additional layers enabled us to generate spectra with increased variability, which was necessary to train our models with the fewest possible spectra and obtain a successful generator. We incorporated a layer named "PoissonNoise" that introduced a Poisson noise distribution into neurons and another layer that applied left and right displacements to them. Our goal with this approach was to simulate the inclusion of random Poisson noise and translations along the energy axis of the generated spectra. It is important to highlight that the number of layers of both Poissonian and Gaussian noise was directly related to the amount of noise in the generated spectra, leading to a poorer signal-to-noise ratio (SNR). We also observed that introducing these layers, *PoissonNoise*, *GaussianNoise*, and *Translation*, in a more disorganized, nonsystematic way resulted in spectra with significantly more variance than when they were systematically incorporated into the generator architecture.

Regarding the Poisson noise added during this work in different parts of the DAG strategy, it is important to note that

for the modern detectors, i.e., the hybrid pixel direct electron counting detectors, adding Poisson noise is actually very realistic. These detectors record very little noise, and this noise follows an almost perfectly Poisson distribution.

Data Augmentation Validation

To evaluate the synthetic spectra produced by our generators, we implemented three validation strategies. The first strategy is automatically executed during training. As mentioned earlier, an initial assessment of the generated spectra is conducted during the training phase by the stopping criteria strategy. Consequently, the generated spectra should at least be compatible with these three metrics of the real spectra. This evaluation is then followed by computing the Kullback–Leibler (KL) divergence as the fourth metric. In addition to these four metrics, we incorporated a fifth, the SNR, which quantifies signal clarity by comparing the desired signal level to the level of background noise. A higher SNR signifies a cleaner signal, thereby enhancing the accuracy of data interpretation and analysis.

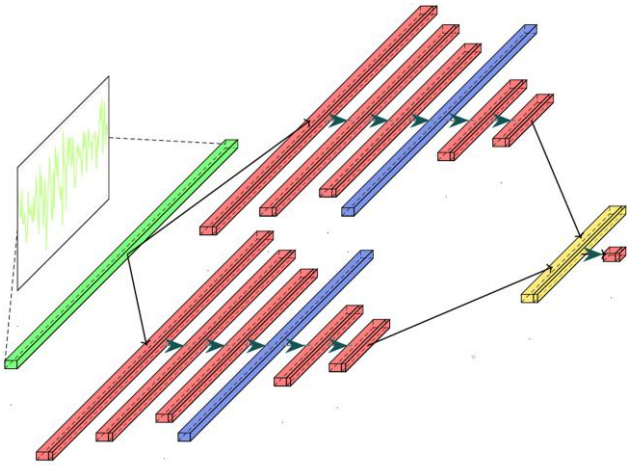


Fig. 4. Schematic of the SimpleDense discriminator's architecture. Each bar represents a layer, and its color indicates the layer's type. The red bar represents dense layers activated by the ReLU function, the blue bars represent the minibatch discriminator layers, and the yellow bar represents a concatenation layer. Note that the last or exit dense layer uses the sigmoid activation function.

However, the use of these five metrics alone does not ensure the generation of high-quality spectra. Additionally, to the use of these five metrics, the quality of the spectra was qualitatively assessed in two visual ways: first, by plotting two individual generated spectra alongside two random spectra from the training data set, not necessarily from the specific subset used for training, and second, by drawing a 2D map of signal intensity versus energy loss, which enables us to visualize multiple spectra concurrently while comparing the generated noise distribution, positions relative to the energy axis, and energy shifts in relation to the expected onset observed in the real spectra. Finally, a comparison was also analyzed in the Supplementary Material to contrast this DA strategy with a more conventional approach, the performance accuracies of which are presented in Supplementary Table S3. This approach involves simply adding noise to the spectra, as shown in the noise-augmented spectra in Supplementary Figures S3 and S4.

Usually, when validating GAN generators, the generated data are evaluated using classifiers previously trained with real data. However, given our primary objective of using the generators as a DA strategy for classifying EEL spectra, we found it more convenient to train the classifiers with synthetic data and then evaluate them on real data. This approach helped to assess the capacity of synthetic data to generalize well to real-world data, thereby providing a more precise evaluation of the performance of the generator.

As classifiers, we adopted two well-established ML strategies for EELS: SVM and ANN classifiers (Chatzidakis & Botton, 2019; del-Pozo-Bueno et al., 2023). The robustness of these classification strategies in front of the very same spectral features as those of the present study has been rigorously demonstrated in previous works of del-Pozo-Bueno et al. (2023). For the soft-margin SVM, we used the radial basis function as a kernel function and optimized the parameters for each data set using Gridsearch, specifically the regularization parameter C and the kernel parameter gamma. As for the ANNs, we used a network with two branches, one convolutional and one dense, which have demonstrated strong performance in classifying EEL spectra.

As the main goal then, the next section starts with the results of accuracy tests of these SVM and ANN classifiers trained with augmented data and continues with the direct comparison between generated and experimental data. In the second

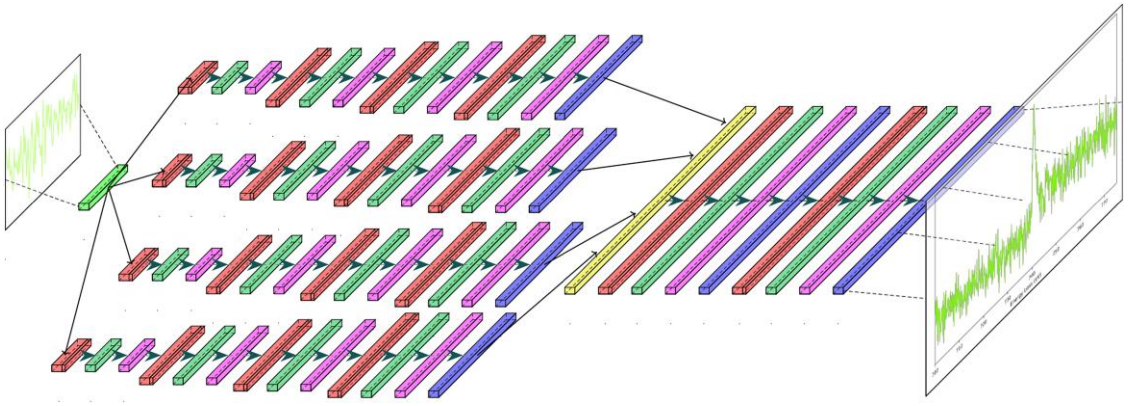


Fig. 5. Schematic of the SimpleDense generator's architecture. Each bar represents a layer, and its color indicates the layer's type. The red bar represents dense layers activated by the ReLU function, the blue bar represents the Translation layers, the purple bar represents the GaussianNoise layers, the green bar represents the PoissonNoise layers, and the yellow bar represents a concatenation layer.

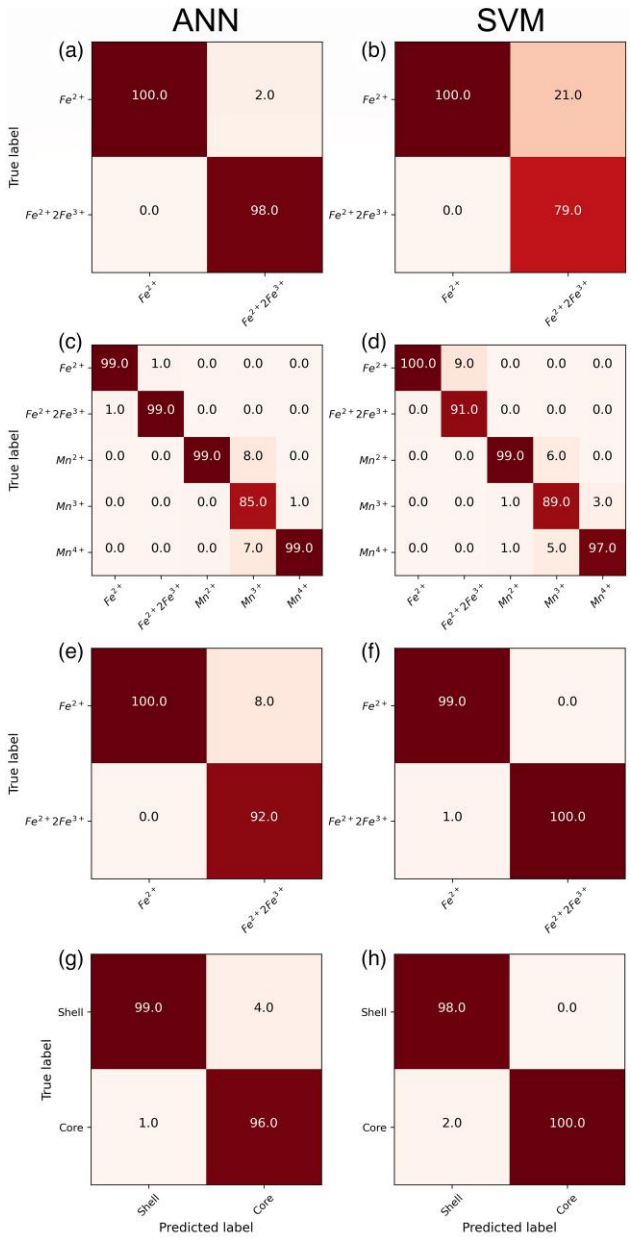


Fig. 6. Confusion matrices for ANN (first column) and SVM (second column) classifiers. These classifiers were trained on synthetic spectra and tested on real ones. Each row of the figure corresponds to a different data set in the order of W, WL, K, and MF from top to bottom. The confusion matrices are column normalized. The models from a to d of the figure were trained with data generated by the Single-BiTrans GAN, while the models from e to h were trained with data from the Multi-BiTrans GAN, as they performed the best in the accuracy test.

part, the major issues encountered during the training of the DAGs and the solutions implemented to overcome them are discussed.

Results and Discussion: Evaluation of GAN Models and Generator Performance

Spectra Generation Evaluation

In this section, we evaluate the spectra generated by our DAG designs. As previously described, four DAG designs were proposed: the Single-MonoTrans and Single-BiTrans GAN models, only containing a single discriminator, and the Multi-MonoTrans and Multi-BiTrans GAN design, containing two and three discriminators, respectively. The first criterion used to determine which of the four models generated better spectra was based on evaluating which of the classifiers, trained with synthetic spectra, classified the real ones more satisfactorily. This evaluation was performed on the four data sets constructed from experimental EELS spectra. Note that less than 15% of the spectra were used for training the DAG models. Therefore, test accuracy of the two classifiers, ANN and SVM, trained with the augmented data generated using the four DAG designs and three distinct architectures, was measured for the different data sets, W, WL, K, and MF. Table 1 presents the test accuracy of ANN and SVM classifiers across various datasets (W, WL, K, and MF), utilizing different DAG designs and architectures. Each cell displays the test accuracy percentage corresponding to each combination of dataset, DAG design, and architecture. The accuracy values for ANN and SVM are placed side by side. They are color-coded with green indicating the best performance and red signifying the worst. Therefore, Table 1 presents an overview of the classification results, allowing for a quick visualization of which DAGs were more efficient according to the color code employed for the best (green tones) and not-so-good (red tones) performances.

The analysis of ML classification in Table 1 revealed significant variability in the performance of the models depending on the data set and the specific DAG design employed. In the W data set, the ANN achieved the highest test accuracy when trained with the Multi-BiTrans GAN, while the SVM performed optimally with the Single-BiTrans GAN. For the WL data set, the Single-BiTrans GAN yielded the highest test accuracy for the ANN, while the Multi-MonoTrans GAN resulted in the best performance for the SVM, followed closely by the Single-BiTrans GAN. In the case of the K data set, both the ANN and SVM classifiers performed best with the Multi-BiTrans GAN. Finally, in the MF data set, the ANN trained with the Single-BiTrans GAN achieved the highest accuracy, while the SVM reached its peak performance when trained with the Single-BiTrans and the Single-MonoTrans GAN. Overall, the SimpleDense and SimpleConvo architectures tended to yield higher test accuracies across all data sets and DAG designs for both classifiers, with the former showing slightly better results. On the other hand, the best performing DAG varied depending on the data set, suggesting that the choice of DA strategy should be data set specific. In general, the Single-BiTrans and Multi-BiTrans GANs appeared to be very effective for the ANN classifiers across multiple data sets, though not exclusively the most effective in every case. Similarly, the best DAG for the SVM varied across data sets.

From these results, we observed that designs with two transformations, translation and noise, performed better than those with only one transformation. In general, both the Single-BiTrans and Multi-BiTrans GANs showed better classification results. Additionally, considering that the convergence for single discriminator models is achieved more rapidly and they exhibit greater stability, the single discriminators prove to be the more advantageous choice, as will be demonstrated in Section DAG convergence evaluation.

A close examination of the efficiency of the classifiers with respect to the different classes is very well visualized by their confusion matrices, as depicted in Figure 6, corresponding to the four W, WL, K, and MF data sets, respectively. The first

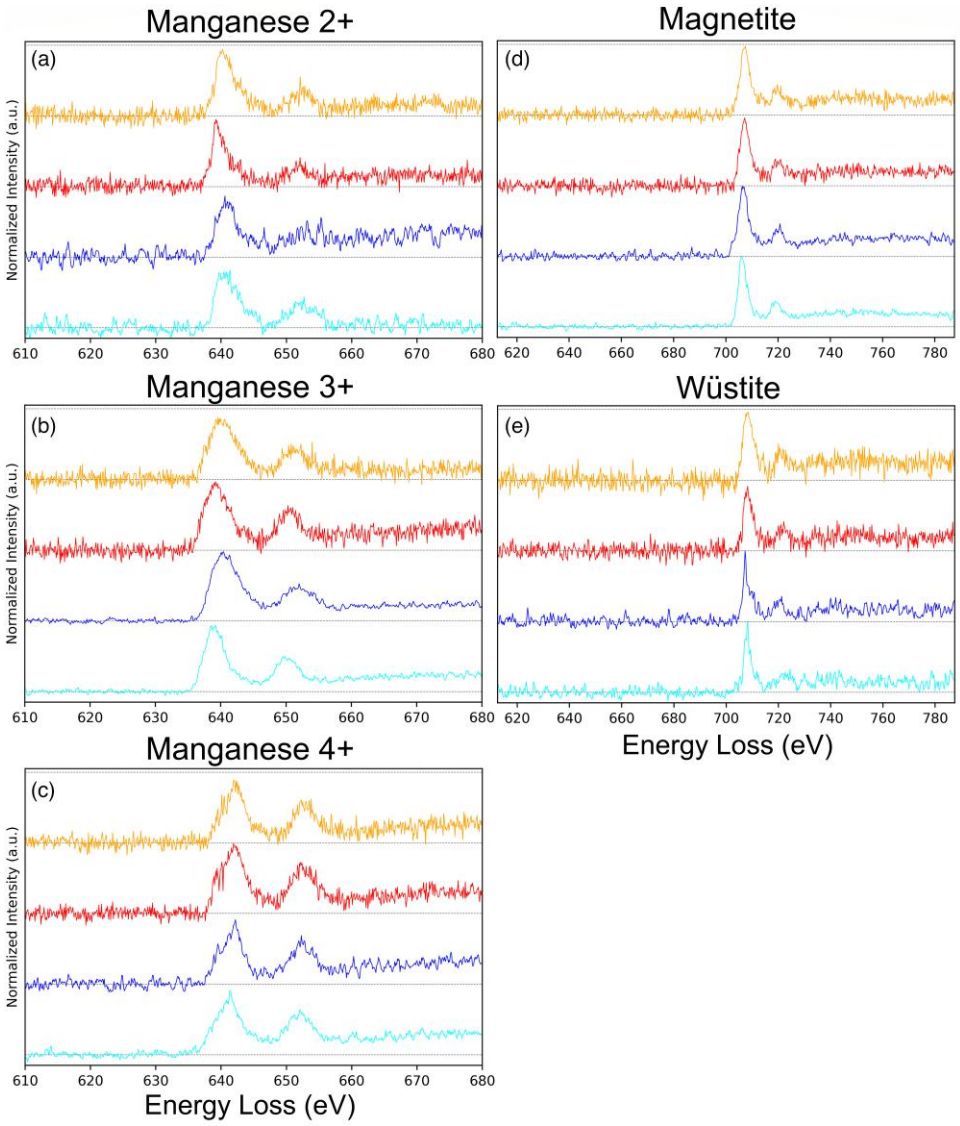


Fig. 7. Real and synthetic EEL spectra for Mn and Fe white lines. The energy range for the Mn white lines spans from 610 to 680 eV, while for the Fe white lines, it covers from 615 to 785 eV. The original spectra are represented by blue and cyan lines, while the generated spectra are depicted in orange and red. The white lines are arranged as follows: (**a**) Mn^{2+}, (**b**) Mn^{3+}, (**c**) Mn^{4+}, (**d**) wüstite, and (**e**) magnetite. These synthetic spectra are generated by the Single-BiTrans GAN design and SimpleDense architecture.

column of the figure represents the ANN classifier, while the second one represents the SVM classifier. It is important to note that we present only the confusion matrices of the best classifiers, which correspond to the Single-BiTrans and Multi-BiTrans GANs using the SimpleDense architecture, where the data for Figures 6a to 6d were generated by the Single-BiTrans GAN and the data for Figures 6e to 6h by the Multi-BiTrans GAN, which achieved better accuracy test results. These confusion matrices illustrate the good performance of the classifiers, with the majority of misclassifications occurring between spectra of the same compound, that is, some misclassifications between different Fe or Mn species but telling apart Fe and Mn white lines.

After analyzing that synthetic spectra could be useful as DA, we further evaluated the generated spectra to confirm their coherence with real ones. For this task, we focused on evaluating the spectra generated by the Single-BiTrans design using the SimpleDense architecture, as it had shown one of the best classification results for all data sets. The evaluation started with the WL data set and then proceeded with the rest.

To evaluate the performance of the Mn and Fe white line generators, Figure 7 presents a visual comparison of individual synthetic spectra with real ones. Each pair of white lines is represented by a total of four spectra, plotted as normalized intensity versus energy loss, with the synthetic spectra shown in orange–red and the real spectra in blue–cyan. Furthermore, Figure 8 displays the spectra as 2D maps, simulating how they would be observed by a spectrometer camera before binning the data in the nondispersive direction. This figure presents the spectra for all the white lines as 2D intensity maps, with the x axis representing the energy loss in electronvolts and each y coordinate corresponding to a single spectrum. These maps show a total of 150 spectra, separated into two groups of 75 spectra. The upper 75 spectra belong to a random subset of the original spectra, while the lower 75 spectra belong to the synthetic spectra.

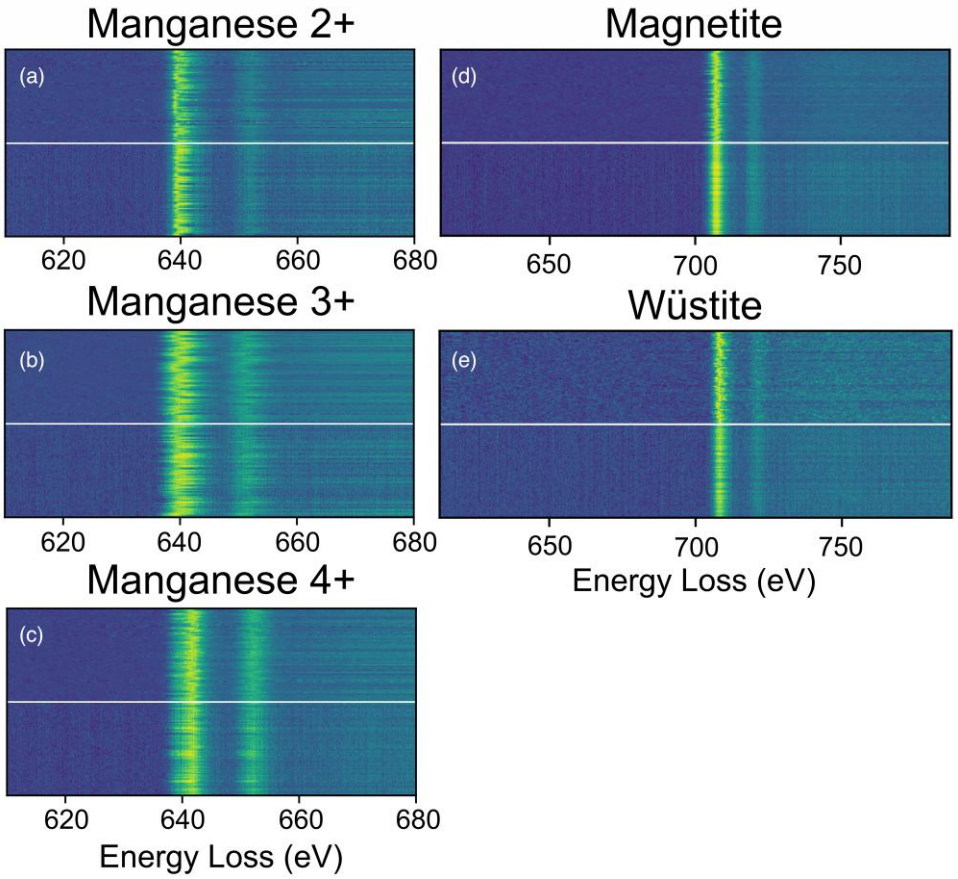


Fig. 8. 2D map representing the real and synthetic spectra of the Mn and Fe white lines, with the generated spectra obtained from the Single-BiTrans GAN design and SimpleDense architecture. The *y* axis represents the individual spectra, while the *x* axis corresponds to the energy axis, spanning from 610 to 680 eV for Mn and from 615 to 785 eV for Fe. A white line divides the map into two sections, where the upper section corresponds to the real spectra, and the lower section corresponds to the generated spectra.

The visual comparison in Figures 7 and 8 demonstrates that the generators accurately replicate the ELNES, i.e., the shape of the edges of the white lines. The synthetic spectra closely match the intensity of the normalized real ones. Looking at the noise distributions, the generated spectra generally exhibit a higher level of noise, compared to the original ones, also presenting a high variability and heterogeneity. Additionally, Figure 8 provides a more detailed examination of the edge onset positions and a larger quantity of spectra in a single view. The variability of the edge onset position and amplitude was accurately reproduced in all cases. Moreover, it was observed that there were small variations in the absolute positions of the white lines, which is a common phenomenon when multiple samples of the same material are visualized together. The generators could reproduce this effect, making it difficult to visually distinguish between real and synthetic spectra. Overall, the visual comparison presented in these figures demonstrates that the spectra generated were realistic and appropriate for human evaluation.

Having evaluated the results for the WL data set, which contained the Fe and Mn white lines separately, we also visually evaluated the spectra for the other data sets that had different features and more complex signals. Figure 9, for the W and K data sets, and Figure 10, for the MF data set, both display the results in a similar manner, providing a visual comparison of individual synthetic spectra with real ones and between the corresponding 2D intensity maps. These figures confirm that these DAG strategies could produce generators capable of producing feasible spectra for different and complex EELS characteristics.

Another relevant aspect to quantify is the SNR of the generated spectra in comparison to the experimental ones. The results are presented in Table 2, which includes the SNR values for the original spectra (first SNR column) and the values for the spectra generated using the different DAG strategies and architectures. Additionally, the table is color coded: in yellow are the values close to the original SNR; red for lower SNR values than the original, indicating higher noise levels; and green for higher SNR values, suggesting lower noise levels. As suggested by the individual spectra in Figure 7, the SNRs of the synthetic spectra for the Single-BiTrans GAN were lower or almost equal to the SNR of the original ones (red or yellow colored).

Table 2 presents the SNR analysis for generated spectra, SNR values for both the orginal and synthetic spectra across all datasets and classes, generated by the various DAG and architectures. These values are color-coded: yellow indicates values close to the orginal SNR, red denotes lower values and green signifies higher values. This Table 2 shows that our models effectively produce spectra with SNRs similar to that of the original spectra, adapting well to the SNR of the training data. These models replicated low SNRs well but struggled with very high SNRs. The architecture or training data set was key in modulating noise levels, with SimpleDense

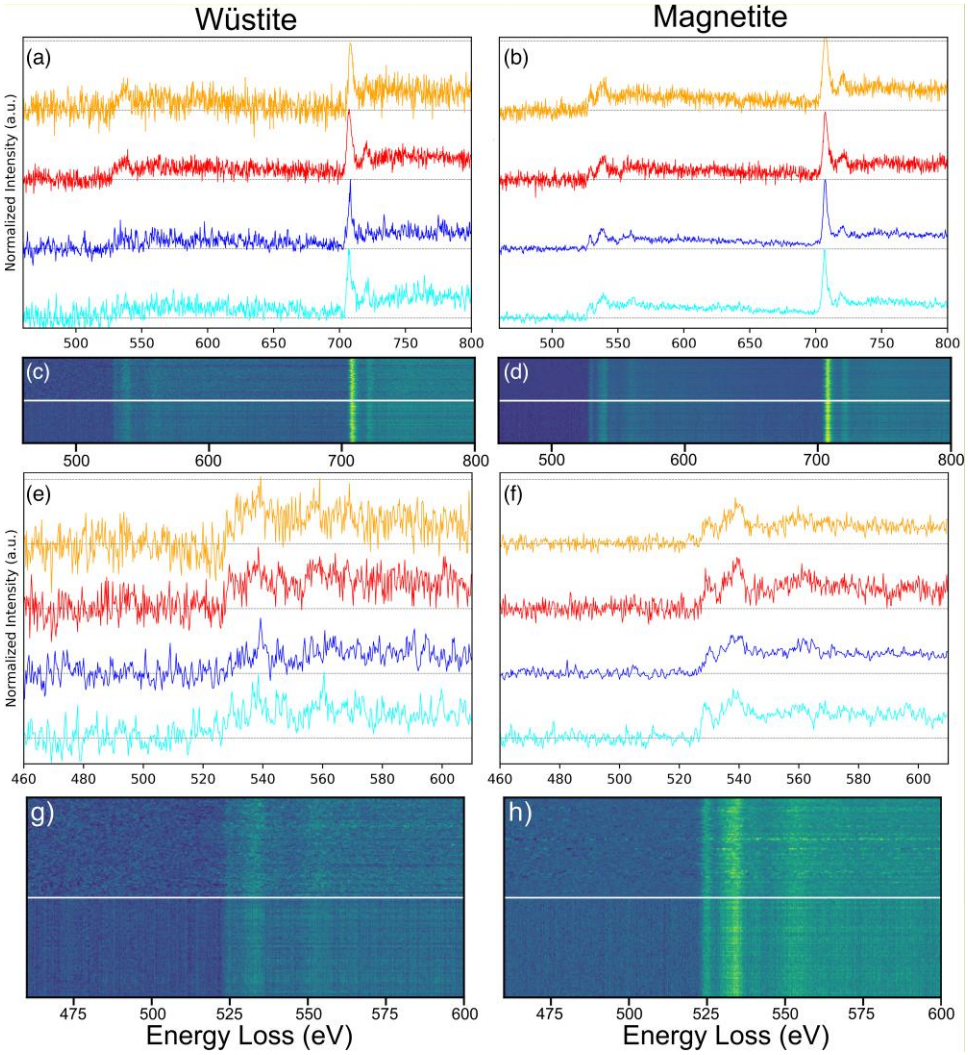


Fig. 9. Comparison of real and synthetic spectra for both the oxygen K-edge and Fe white lines, with an energy range from 460 to 800 eV. All spectra were generated using the Single-BiTrans GAN design and SimpleDense architecture. (**a**) and (**b**) display individual line plots of real and synthetic spectra, while (**c**) and (**d**) depict 2D intensity maps, all for wüstite and magnetite, respectively. (**e**) and (**f**) illustrate individual line plots, similar to those in (**a**) and (**b**), but specifically focused on the oxygen K-edge with an energy range from 460 to 600 eV. (**g**) and (**h**) provide 2D intensity maps analogous to (**c**) and (**d**), all for wüstite and magnetite, respectively.

outperforming LargeDense and SimpleConvo. Generally, SimpleDense generated noisier data, while SimpleConvo occasionally produced cleaner spectra than the original. In this regard, all architectures incorporated layers of Poisson and Gaussian noise, thereby creating a baseline noise level and upper bounding the SNR and ensuring a minimum noise level in the spectra produced by these architectures.

These generative models were sensitive to the training data set, with inconsistencies in replicating SNRs, especially in composites with low noise levels like magnetite in the W and MF data sets. Comparing SimpleDense and LargeDense in the Single-MonoTrans and Multi-MonoTrans GANs, the latter showed increased SNR, indicating less noise due to fewer noise layers, thus suggesting control over noise levels in spectra by adjusting noise layers. Nonetheless, for DA purposes, lower SNR is preferable, as shown in Table 1, to train classifiers on varied spectra and prevent overfitting. In summary, the generators were effective in replicating the noise levels of the training spectra, constrained always by the inherent noise embedded within the architecture.

Data Augmentation Generative Adversarial Network Convergence Evaluation

In this last section, we evaluated the performance of the training setup for the DAG models by examining their metrics and costs or losses while discussing around the convergence patterns observed and the nuances of the early stopping strategy implemented. Initially, we analyzed the cost, assessing the cost values of the discriminator and the generator for each training epoch. It should be noted that the cost function used in our models corresponds to binary cross-entropy. Ideally, in a GAN model, the generator and discriminator costs would converge to a point where the generator produces perfectly realistic images. However, achieving this balance in practice can be challenging, and fluctuations and oscillations in the costs are common (Salimans et al., 2016; Heusel et al., 2017; Brock et al., 2018).

Figure 11 presents the costs as a function of the epoch for all DAG designs using the SimpleDense architecture trained on the MF data set, specifically for the class corresponding to Mn–

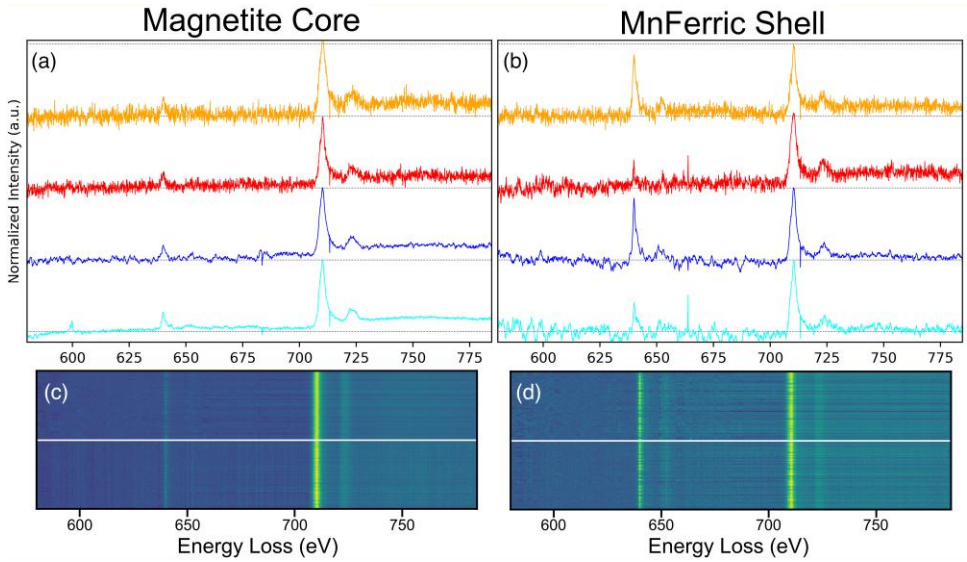


Fig. 10. Comparison of real and synthetic spectra for both Fe and Mn white lines, with an energy range from 580 to 784.7 eV. All spectra were generated using Single-BiTrans GAN design and SimpleDense architecture. (**a**) and (**b**) display individual line plots of real and synthetic spectra, while (**c**) and (**d**) depict 2D intensity maps, all for magnetite core and Mn–ferrite shell, respectively.

Table 2. "Signal-to-Noise Ratio (SNR) Analysis for Generated Spectra." SNR values for both the original and synthetic spectra across all datasets and classes, generated by the various DAGs and architectures. The SNR values are color-coded: yellow indicates values close to the original SNR, red denotes lower values, and green signifies higher values.

Dataset	Composites	Signal to Noise Ratio (SNR)									
		Original	Single-MonoTrans			Multi-MonoTrans			Single-BiTrans		Multi-BiTrans	
			Simple Dense	Large Dense	Simple Convo	Simple Dense	Large Dense	Simple Convo	Simple Dense	Simple Convo	Simple Dense	Simple Convo
WL	Wüstite (Fe^{2+})	1,1	1,4	1,7	1,2	1,4	1,8	1,3	1,2	1,4	1,3	1,2
	Magnetite (Fe^{2+} $2Fe^{3+}$)	2,1	1,6	2,6	2,1	2,0	2,1	2,5	2,1	2,7	2,0	2,6
	Mn^{2+}	1,5	1,4	2,0	1,2	1,5	1,8	2,7	1,4	2,1	1,5	2,3
	Mn^{3+}	2,2	2,0	1,9	2,4	1,8	2,5	2,5	1,7	2,5	1,6	3,4
	Mn^{4+}	2,5	1,9	2,8	2,8	2,6	2,6	3,5	2,1	2,8	2,1	3,9
W	Wüstite (Fe^{2+})	1,6	2,1	2,8	2,0	2,0	2,8	2,0	1,7	1,6	1,7	2,3
	Magnetite (Fe^{2+} $2Fe^{3+}$)	3,3	2,6	3,3	4,2	2,5	2,8	4,6	2,1	6,9	2,2	5,0
K	Wüstite (Fe^{2+})	0,9	1,0	1,0	1,0	1,0	1,0	1,1	1,2	1,1	1,2	1,6
	Magnetite (Fe^{2+} $2Fe^{3+}$)	2,2	2,1	2,0	3,0	2,3	2,0	2,3	2,1	4,8	2,1	3,3
MF	Magnetite Core	2,8	1,4	1,6	2,2	1,3	1,6	2,1	1,6	3,3	1,4	2,7
	Mn-Ferric Shell	1,5	1,5	2,4	1,5	1,4	1,8	1,7	1,4	1,4	1,7	1,8

ferric shell. The cost axis is on a logarithmic scale, as this facilitates the visualization and comparison of the different cost values. From these figures and the costs of the other models, we observed that, for the discriminators, almost all models reduced their costs during training. Additionally, Figure 11 illustrates that both the generator and the discriminator did not seem to converge to a specific value; rather, they fluctuated and oscillated throughout the entire training process. This suggests that the convergence of the models may not be still completely achieved since we have applied an early stopping strategy. It should be noted that in Figures 11a and 11c, some anomalies such as abrupt and sharp peaks are observed. These anomalies, consisting of an increase in the cost by both the generator and the discriminator, could be attributed to multiple factors. First, it indicates instability during the training of our GANs, suggesting difficulties in converging toward an equilibrium point, i.e., the GAN is struggling to find a stable and satisfactory solution. Another possible reason could be related to exploding gradients, where very large gradients cause abrupt increases in the costs of both networks. However, despite the lack of apparent convergence of the generator and discriminator costs, both models show improvement and evolution in their task of generating realistic spectra. Therefore, despite the lack of an apparent robust convergence of the generator and discriminator costs,

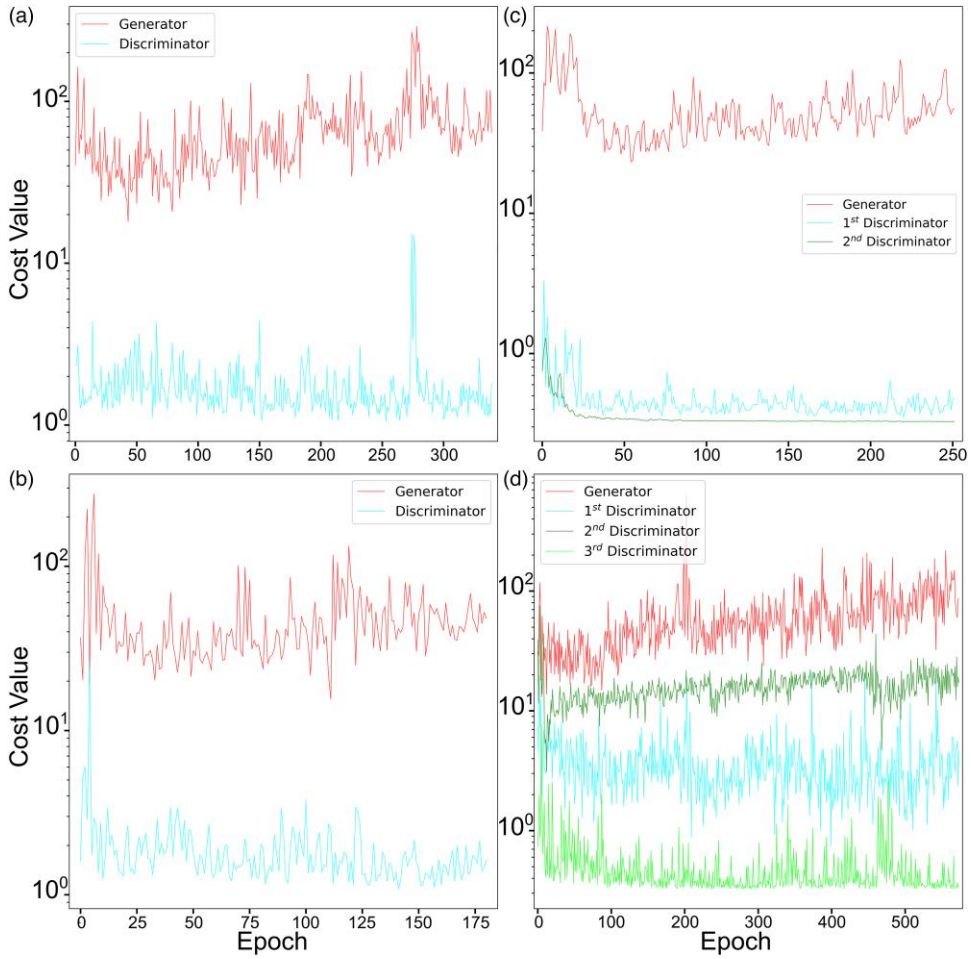


Fig. 11. Cost trends for the four DAG designs within the SimpleDense architecture, specifically trained on the MF data set corresponding to the Mn–ferric shell class. (**a**)–(**d**) represent the Single-MonoTrans, Multi-MonoTrans, Single-BiTrans, and Multi-BiTrans GAN, respectively. The costs are plotted against epochs on a logarithmic scale to facilitate straightforward comparisons.

our models performed their tasks properly. It is crucial to note that successful training does not necessarily entail the convergence of the generator and discriminator costs. Some GANs can yield high-quality results even if the costs do not appear to stabilize (Heusel et al., 2017).

In contrast to Figure 11, where training was early stopped, in Figure 12 (Supplementary Figure S5), we present the cost values for models trained up to 7,500 epochs to ensure their convergence. As in the previous graphs, the costs of both the generator and the discriminator exhibit fluctuations, which is a common phenomenon during GAN training, as they are engaged in a kind of "game," competing against each other (Goodfellow et al., 2014). Despite these fluctuations, for single discriminator models, the losses stabilize around 1,000 epochs. For models with multiple discriminators, as shown in graphs b) and d), the generators initially face significant loss but stabilize quite rapidly over time. In graph b), the first discriminator stabilizes fast, while the second shows varied behavior, hinting at challenges in differentiating real from generated data. Graph d) reflects a similar trend: the first and third discriminators stabilize quickly, but the second has noticeable fluctuations. Across the board, all components appear to find a balance. Overall, single discriminator models seem more stable than multidiscriminator ones, with BiTrans models showing strong stability.

In light of these cost training results and the early stopping criteria, we affirm that this strategy is well grounded. First, it assists in maintaining the model generalizable and avoids excessive adaptation to the training data. Furthermore, it leads to a significant reduction in resource consumption (as seen in Supplementary Table S4), optimizing both time and computational power. Crucially, this choice to adopt early stopping was largely driven by the capability of the DAG models to generate convincingly realistic spectra (as seen in Supplementary Figures S6 and S7), thereby ensuring the primary work objectives were met (as seen in Supplementary Table S5). A deeper discussion into the effectiveness of our early stopping strategy is provided in the Supplementary Material.

To complement this, we computed the evaluation metrics during training. Figure 13 displays the training metrics as a function of the epochs. Figures 13a and 13c illustrate the FID, while Figures 13b and 13d present both the PCC and the CosD, given their similar range of values. These plots revealed that both DAG models were converging properly as they improved the generation throughout the epochs. They also met the stopping criterion after a relatively small number of iterations. During the initial epochs, the stopping metrics significantly reduced. Following this rapid approach to the stopping metrics, the final values were gradually achieved over a smooth slope. It is worth noting that the number of

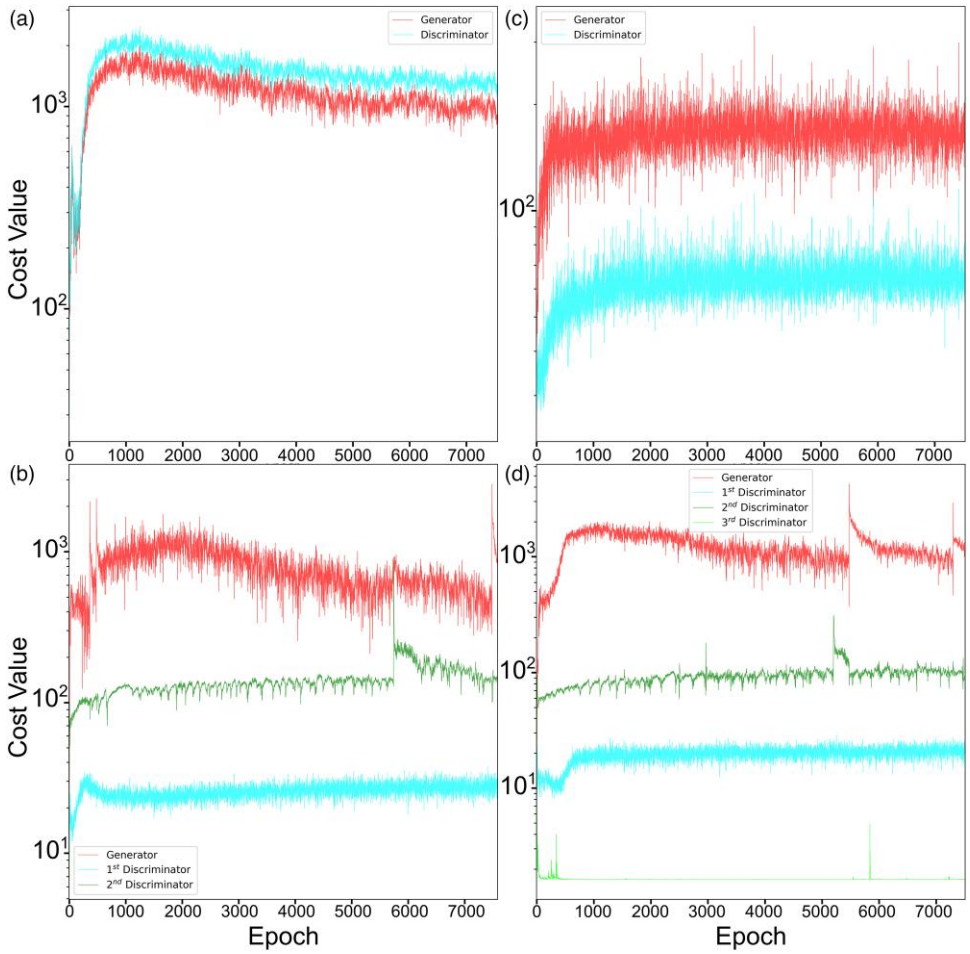


Fig. 12. Cost trends for the four DAG designs within the SimpleDense architecture, specifically trained on the MF data set corresponding to the Mn–ferric shell class. These GANs were trained up to 7,500 epochs to ensure their convergence. (**a**)–(**d**) represent the Single-MonoTrans, Multi-MonoTrans, Single-BiTrans, and Multi-BiTrans GANs, respectively. The costs are plotted against epochs on a logarithmic scale to facilitate straightforward comparisons.

epochs required for convergence varies depending on the complexity of the spectra being generated, in our case the number of energy channels. Therefore, the successful generators were obtained faster for the WL data set than for the MF data set. Additionally, the number of epochs for a successful generator was influenced by the number of discriminators to be trained. Thus, generally, the Multi-MonoTrans and Multi-BiTrans GANs required more epochs to achieve the same level of realism as the Single-MonoTrans and Single-BiTrans GAN designs.

Regarding the KL divergence, which is evaluated *a posteriori* and not during training, it shows that the generated spectra were compatible with the experimental spectra in accordance with the other three metrics. These results are presented and discussed in the Supplementary Material, in particular Supplementary Table S2.

Overall, these results and analysis presented in this section highlight both the performance and challenges faced during the evaluation of GAN models for generating synthetic EEL spectra. The proposed models effectively addressed major issues, such as mode collapse, through techniques like mini-batch discrimination and the inclusion of noise layers. While oscillations were observed in the costs throughout training, evaluation metrics demonstrated promising progress and convergence, underscoring the capability of GANs to generate realistic spectra. In this sense, the early stopping strategy proved to be well founded, conserving resources while ensuring the generation of realistic spectra. Therefore, these findings contribute to the understanding and improvement of GAN performance in generating realistic spectra.

In summary, results presented in this work demonstrate that the Single-BiTrans GAN and Multi-BiTrans GAN designs were capable of generating more realistic synthetic EELS spectra than the Single-MonoTrans GAN and Multi-MonoTrans GAN models, proving that a larger number of transformations were more convenient. In addition, regarding the convergence of the models, it was observed that the single models were more stable and presented a faster convergence; thus, they were more convenient. These synthetic spectra were effectively used for DA and successfully trained ML classifiers. The superior performance of the Single-BiTrans GAN and Multi-BiTrans GAN designs was demonstrated by their higher accuracy in classifying real spectra, as well as their closer similarity to the original spectra as measured by FID, COR, CosD, and KL divergence. Additionally, since stopping criteria were included in the training scheme, our DAGs required fewer training iterations and fewer computational resources compared to other GAN or DAG schemes. These findings confirm that DAG and, of course, GANs were effective in generating EELS spectra for classification purposes.

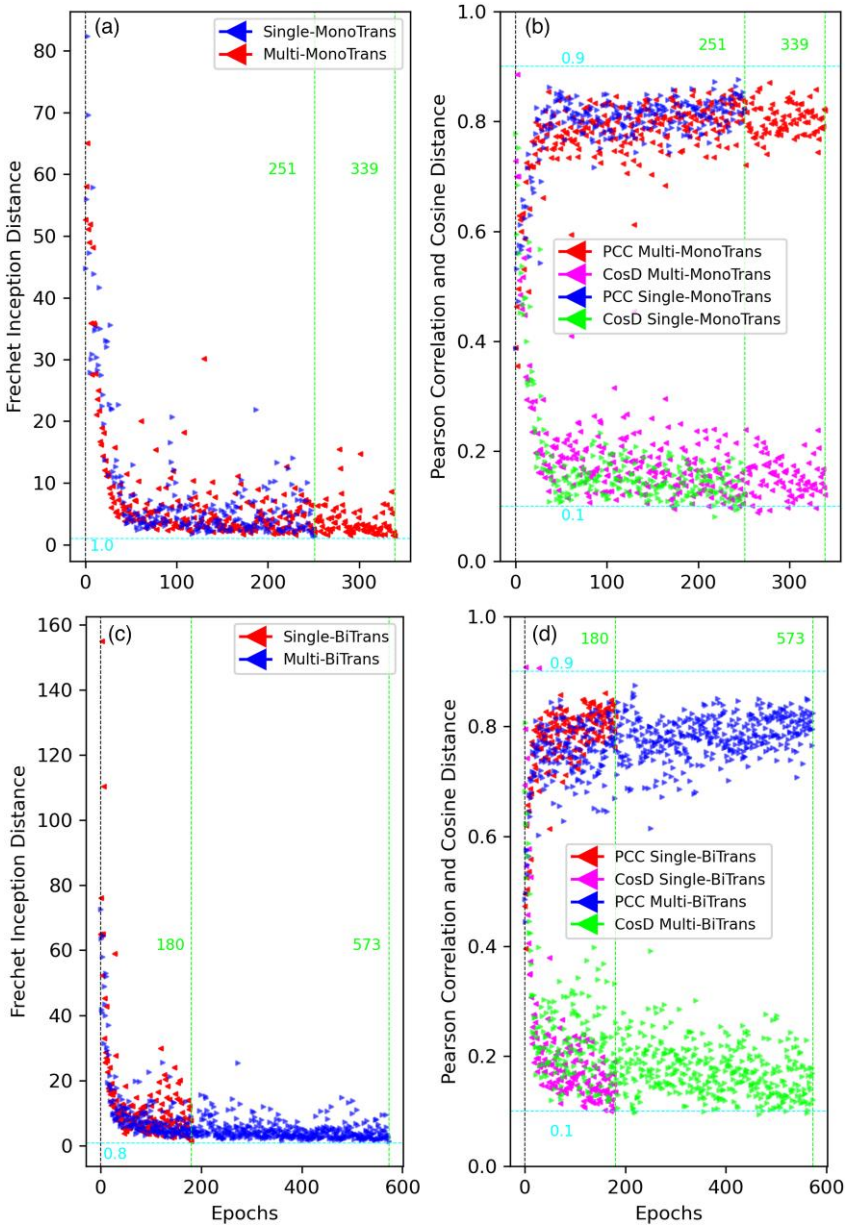


Fig. 13. Trend plots of the training metrics for the four DAG designs using the SimpleDense architecture, specifically trained on the MF data set corresponding to the Mn–ferric shell class. The Fréchet inception distance (FID) is plotted individually, while the Pearson correlation coefficient (PCC) and cosine distance (CosD) are plotted together due to their similar value ranges. (**a**) and (**b**) represent the FID for the Single/Multi-MonoTrans GANs and Single/Multi-BiTrans GANs, respectively, while (**c**) and (**d**) represent PCC and CosD for the Single/Multi-MonoTrans GANs and Single/Multi-BiTrans GANs, respectively.

Finally, it is important to emphasize that the primary role of the DAG strategy in this research is as a DA technique. As we use GANs for this task, it is important to understand that they parametrize the information inherent in the original spectra to produce new ones. However, they cannot generate genuinely new information or features; they just combine the existing characteristics. Consequently, our objective is to generate an extensive and varied set of training data from, as much as possible, diverse experimental spectra, ensuring that the synthesized data not only retain but also combine the features present in the original spectra. A major advantage of this approach is its ability to combine characteristics of spectra from disparate experiments and samples, enhancing the richness of our data set. Notably, GANs can mix the attributes of these spectra, yielding outputs that incorporate features from a range of samples and microscopes, thus boosting the dataset's diversity and representativeness.

Conclusions

This study demonstrates that GANs are an effective strategy for generating normalized EELS spectra. Furthermore, the generators produced by these GANs work as a DA strategy for training supervised or semisupervised ML algorithms, such as SVMs and ANNs. Moreover, we demonstrate that integrating the DA strategy for GANs, known as DAG, allows for training with a reduced set of experimental EEL spectra while simultaneously maintaining the high quality of the generated spectra.

Regarding the results from the DAG designs, our analysis concludes that even with just two transformations with physical meaning, regardless of the number of discriminators, we can generate high-quality spectra rapidly and efficiently. The most effective DAG designs for DA correspond to the Single-BiTrans and Multi-BiTrans designs, those encompassing two transformations: random translations along the energy–loss axis and random incorporation of Poisson noise. This study indicates that models with fewer discriminators lead to more stable and convergent models. Therefore, the Single design, which employs a single discriminator, is generally preferable to the Multi design, which uses three. Despite this, the performance of Multi models is not significantly inferior. Both the Single-BiTrans and Multi-BiTrans designs, despite their difference in the number of discriminators, succeed in generating high-quality spectra. However, a higher number of discriminators can potentially slow down the convergence of our training.

An essential aspect of our work is the implementation of a stopping mechanism in the GAN training. It is important to emphasize that while this mechanism does not always guarantee the stable convergence of the GAN when training is halted, it does ensure the quality of the generated data. This approach significantly reduces the computational cost of the training while still achieving high-quality generators.

Lastly, this DAG approach enables the combination of spectra from multiple experiments and samples, thereby significantly enhancing the diversity of the generated data beyond the need for large volumes of experimental spectra. Since GANs utilize existing data distributions to synthesize new spectra by recombining, not creating new, features from the training data, the diversity of data, rather than its quantity, should guide the training of DAG models.

Availability of Data and Materials

The authors have declared that no data sets apply for this piece.

Supplementary Material

To view supplementary material for this article, please visit https://doi.org/10.1093/mam/ozae014.

Acknowledgments

The authors would like to thank Prof. Vlado Lazarov (University of York) and Prof. Sara A. Majetich (Carnegie Mellon University) for the use of data from Mn–ferrite NP samples; and DP would like to acknowledge the wise advices from his colleague Ricard Monge Calvo.

Financial Support

This work has been supported by the Spanish Project PDC2021-121366-I00 financed by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. The authors also acknowledge funding from Ministerio de Ciencia, Innovación y Universidades (MCIN) under the projects PID2019-106165GB-C21 and PID2022-138543NB-C21 financed by MCIN/AEI/10.13039/501100011033; the support received from the ELECMI—ICTS Electron Microscopy for Materials Science; the funding from Generalitat de Catalunya under project 2021SGR00242; and the 2020 FI-SDUR 00035 grant from the AGAUR agency of the Generalitat de Catalunya. F. P. acknowledges the ICREA Academia 2022 grant. SuperSTEM is the UK National Research Facility for Advanced Electron Microscopy, funded by the Engineering and Physical Sciences Research Council (EP/W021080/1).

Conflict of Interest

The authors declare that they have no competing interest.

References

Arora S, Ge R, Liang Y, Ma T & Zhang Y (2017). Generalization and equilibrium in generative adversarial nets (GANs). http://arxiv.org/abs/1703.00573.

Blanco-Portals J, Peiró F & Estradé S (2022). Strategies for EELS data analysis. Introducing UMAP and HDBSCAN for dimensionality reduction and clustering. *Microsc Microanal* **28**, 109–122. https://doi.org/10.1017/S1431927621013696

Bonnet N, Brun N & Colliex C (1999). Extracting information from sequences of spatially resolved EELS spectra using multivariate statistical analysis. *Ultramicroscopy* **77**, 97–112. https://doi.org/10.1016/S0304-3991(99)00042-X

Bonnet N & Nuzillard D (2005). Independent component analysis: A new possibility for analysing series of electron energy loss spectra. *Ultramicroscopy* **102**, 327–337. https://doi.org/10.1016/j.ultramic.2004.11.003

Bowles C, Chen L, Guerrero R, Bentley P, Gunn R, Hammers A, Dickie DA, Hernández MV, Wardlaw J & Rueckert D (2018). GAN augmentation: augmenting training data using generative adversarial networks. http://arxiv.org/abs/1810.10863.

Brock A, Donahue J & Simonyan K (2018). Large scale GAN training for high fidelity natural image synthesis. http://arxiv.org/abs/1809.11096.

Chatzidakis M & Botton GA (2019). Towards calibration-invariant spectroscopy using deep learning. *Sci Rep* **9**, 2126. https://doi.org/10.1038/s41598-019-38482-1

Colliex C, Manoubi T & Ortiz C (1991). Electron-energy-loss-spectroscopy near-edge fine structures in the iron-oxygen system. *Phys Rev B* **44**, 11402–11411. https://doi.org/10.1103/PhysRevB.44.11402

del-Pozo-Bueno D, Kepaptsoglou D, Peiró F & Estradé S (2023). Comparative of machine learning classification strategies for electron energy loss spectroscopy: Support vector machines and artificial neural networks. *Ultramicroscopy* **253**, 113828. https://doi.org/10.1016/j.ultramic.2023.113828.

del-Pozo-Bueno D, Peiró F & Estradé S (2021). Support vector machine for EELS oxidation state determination. *Ultramicroscopy* **221**, 113190. https://doi.org/10.1016/j.ultramic.2020.113190.

del-Pozo-Bueno D, Varela M, Estrader M, López-Ortega A, Roca AG, Nogués J, Peiró F & Estradé S (2021). Direct evidence of a graded magnetic interface in bimagnetic core/shell nanoparticles using electron magnetic circular dichroism (EMCD). *Nano Lett* **21**, 6923–6930. https://doi.org/10.1021/acs.nanolett.1c02089

de la Peña F, Berger MH, Hochepied JF, Dynys F, Stephan O & Walls M (2011). Mapping titanium and tin oxide phases using EELS: An application of independent component analysis. *Ultramicroscopy* **111**, 169–176. https://doi.org/10.1016/j.ultramic.2010.10.001

de la Peña F, Prestat E, Fauske VT, Burdet P, Lähnemann J, Jokubauskas P, Furnival T, Nord M, Ostasevicius T, MacArthur KE, Johnstone DN, Sarahan M, Taillon J, Aarholt T; pquinn-dls, Migunov V, Eljarrat A, Caron J, Francis C & Ånes HW (2022). hyperspy/hyperspy: Release v1.7.3 (v1.7.3). https://doi.org/10.5281/zenodo.7263263

Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J & Greenspan H (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **321**, 321–331. https://doi.org/10.1016/j.neucom.2018.09.013

Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A & Bengio Y (2014). Generative adversarial nets. http://www.github.com/goodfeli/adversarial.

Halevy A, Norvig P & Pereira F (2009). The unreasonable effectiveness of data. *IEEE Intell Syst* **24**, 8–12. https://doi.org/10.1109/MIS.2009.36

Heusel M, Ramsauer H, Unterthiner T, Nessler B & Hochreiter S (2017). GANs trained by a two time-scale update rule converge to a local nash equilibrium. http://arxiv.org/abs/1706.08500.

Ishak MA, Abdul Latiff AH, Ho ETW, Fuad MIA, Tan NW, Sajid M & Elsebakhi E (2023). Advanced elastic and reservoir properties prediction through generative adversarial network. *Appl Sci (Switzerland)* **13**(10), 6311. https://doi.org/10.3390/app13106311.

Ji F, Zhang X & Zhao J (2023). α-EGAN: α-energy distance GAN with an early stopping rule. *Comput Vis Image Understanding* **234**, 103748. https://doi.org/10.1016/j.cviu.2023.103748.

Kingma, D. P. & Ba, J. (2014). Adam: a method for stochastic optimization. http://arxiv.org/abs/1412.6980.

Maharana K, Mondal S & Nemade B (2022). A review: Data preprocessing and data augmentation techniques. *Glob Trans Proc* **3**, 91–99. https://doi.org/10.1016/j.gltp.2022.04.020

Muro-Cruces J, Roca AG, López-Ortega A, Fantechi E, del-Pozo-Bueno D, Estradé S, Peiró F, Sepúlveda B, Pineider F, Sangregorio C & Nogues J (2019). Precise size control of the growth of Fe $_3$ O $_4$ nanocubes over a wide size range using a rationally designed one-pot synthesis. *ACS Nano* **13**, 7716–7728. https://doi.org/10.1021/acsnano.9b01281

Oberdick SD, Abdelgawad A, Moya C, Mesbahi-Vasey S, Kepaptsoglou D, Lazarov VK, Evans RFL, Meilak D, Skoropata E, van Lierop J, Hunt-Isaak I, Pan H, Ijiri Y, Krycka KL, Borchers JA & Majetich SA (2018). Spin canting across core/shell $Fe_3O_4/Mn_xFe_{3-x}O_4$ nanoparticles. *Sci Rep* **8**, 3425. https://doi.org/10.1038/s41598-018-21626-0

Pate CM, Hart JL & Taheri ML (2021). RapidEELS: Machine learning for denoising and classification in rapid acquisition electron energy loss spectroscopy. *Sci Rep* **11**, 19515. https://doi.org/10.1038/s41598-021-97668-8

Pelaez-Fernandez M, Majérus B, Funes-Hernando D, Dufour R, Duvail JL, Henrard L & Arenal R (2022). Toward laser-induced tuning of plasmonic response in high aspect ratio gold nanostructures. *Nanophotonics* **11**, 3719–3728. https://doi.org/10.1515/nanoph-2022-0193

Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A & Chen X (2016). Improved techniques for training GANs. http://arxiv.org/abs/1606.03498.

Sandfort V, Yan K, Pickhardt PJ & Summers RM (2019). Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci Rep* **9**, 16884. https://doi.org/10.1038/s41598-019-52737-x

Szegedy C, Vanhoucke V, Ioffe S, Shlens J & Wojna Z (2015). Rethinking the inception architecture for computer vision. http://arxiv.org/abs/1512.00567.

Tan H, Verbeeck J, Abakumov A & Van Tendeloo G (2012). Oxidation state and chemical shift investigation in transition metal oxides by EELS. *Ultramicroscopy* **116**, 24–33. https://doi.org/10.1016/j.ultramic.2012.03.002

Tanaka F H K dos S & Aranha C (2019). Data augmentation using GANs. http://arxiv.org/abs/1904.09135.

Torruella P, Arenal R, de la Peña F, Saghi Z, Yedra L, Eljarrat A, López-Conesa L, Estrader M, López-Ortega A, Salazar-Alvarez G, Nogués J, Ducati C, Midgley PA, Peiró F & Estradé S (2016). 3D visualization of the iron oxidation state in FeO/Fe_3O_4 core–shell nanocubes from electron energy loss tomography. *Nano Lett* **16**, 5068–5073. https://doi.org/10.1021/acs.nanolett.6b01922

Torruella P, Estrader M, López-Ortega A, Baró MD, Varela M, Peiró F & Estradé S (2018). Clustering analysis strategies for electron energy loss spectroscopy (EELS). *Ultramicroscopy* **185**, 42–48. https://doi.org/10.1016/j.ultramic.2017.11.010

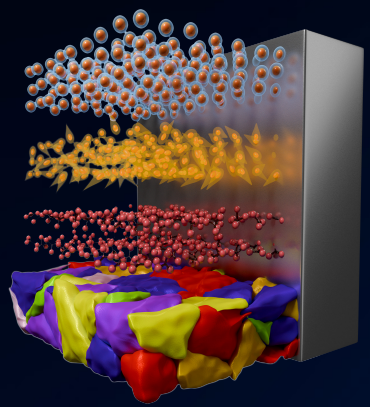
Tran N-T, Tran V-H, Nguyen N-B, Nguyen T-K & Cheung N-M (2021). On data augmentation for GAN training. *IEEE Trans Image Process* **30**, 1882–1897. https://doi.org/10.1109/TIP.2021.3049346

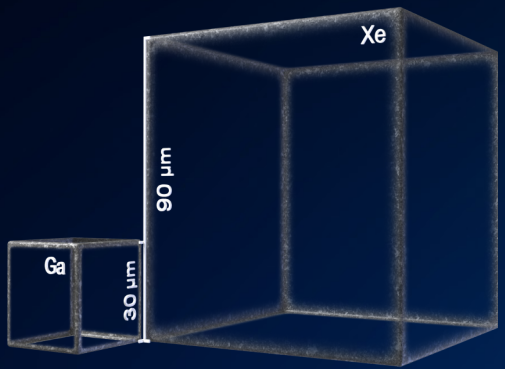
Yedra L, Eljarrat A, Arenal R, Pellicer E, Cabo M, López-Ortega A, Estrader M, Sort J, Baró MD, Estradé S & Peiró F (2012). EEL spectroscopic tomography: Towards a new dimension in nanomaterials analysis. *Ultramicroscopy* **122**, 12–18. https://doi.org/10.1016/j.ultramic.2012.07.020

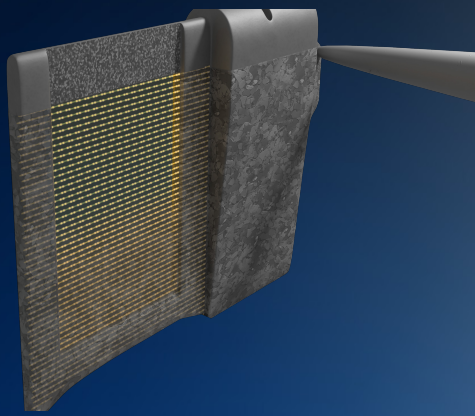
TESCAN
PERFORMANCE IN NANOSPACE

TESCAN AMBER X 2

PLASMA FIB-SEM REDEFINED







UTILITY REDEFINED

SPEED REDEFINED

PRECISION REDEFINED

info.tescan.com