Materials Horizons

NestedAE: Interpretable Nested Autoencoders for Multi-Scale Material Modelling

Journal:	*Materials Horizons*
Manuscript ID	MH-COM-09-2023-001484.R1
Article Type:	Communication
Date Submitted by the Author:	13-Nov-2023
Complete List of Authors:	Thota, Nikhil; Johns Hopkins University, Chemical and Biomolecular Engineering Sharma Priyadarshini, Maitreyee; Johns Hopkins University - Homewood Campus, Chemistry, Chemical and Biolmolecular Engineering; Hernandez, Rigoberto; Johns Hopkins University, Chemistry

Materials Horizons

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

NestedAE: Interpretable Nested Autoencoders for Multi-Scale Material Modeling†

Nikhil Thota,[‡b] Maitreyee Sharma Priyadarshini[‡ab] and Rigoberto Hernandez[*abc]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

We introduce an interpretable machine learning architecture, NestedAE, for multiscale materials using nested supervised autoencoders. We benchmarked the performance of NestedAE on two databases: (1) a synthetic dataset created from nested analytical functions whose dimensionality is therefore known a priori, and (2) a multiscale Metal Halide Perovskite (MHP) dataset that is a combination of an open source dataset containing atomic and ionic properties, and a second dataset containing device characterization using current density-voltage (J-V) analysis. The NestedAE architecture was found to have higher noise robustness and lower reconstruction losses when compared to a vanilla autoencoder (AE). Its application on the MHP dataset revealed links between crystal scale properties and device performance in agreement with earlier experimental observations.

New Concepts

Multiscale modeling and characterization of materials is of paramount importance in ensuring device performance and robustness. The ensuing multivariate figures of merit are proving to be necessary in next-generation technologies capable of addressing global challenges in health, sustainability and energy. In this paper, we introduce a neural network architecture—viz a Nested Autoencoder (NestedAE)—to correlate and predict material properties at many length scales using information from multiple data sources. Each physical scale of the material is represented by an autoencoder. The key contributing factor and novelty of the network architecture is the flow of "important" information from one scale to another through the latent space of each autoencoder. While machine learning techniques have been used to connect the atomic-scale properties of Metal Halide Perovskites (MHPs) with macroscopic device properties, we find that incorporating properties through intermediate (or middle) scales provides a richer description of the materials, and expect that its inclusion will be useful in the future design of these and other materials.

1 Introduction

The need to address the alarmingly high carbon footprint of fossil fuel based energy sources is driving the development of reliable, efficient and economically viable renewable energy sources.[1] Solar cells based on Metal Halide Perovskites (MHPs) are a popular choice for addressing this challenge. Unfortunately, a full characterization of the multiscale behavior of these materials is not yet available, but data science techniques offer a promising route enabling design with or without it.

In general, multiscale materials modeling has become a vital tool in the discovery of functional materials, the characterization of the physical processes governing material behavior, and the design of materials with targeted properties across different length scales.[2–5] Physics-based (or bottom-up) approaches are typically built up from well-understood models at the finest (or finer) scales through exact or approximate coarse-graining methods.[6–8] Top-down approaches typically rely on the identification of coarse-grained variables interacting through effective—that is, optimized or semi-empirical—equations of motion to describe the behavior at a given scale and infer the necessary properties of the materials at the finer-grained scales.[9,10] Here, we provide one answer to the question of whether machine learning techniques can address such hierarchies of scales in a complex material system without requiring an *a priori* assumption about the direction of the scales—e.g., top-down or bottom-up.

Machine learning algorithms have been used to predict many properties of MHP materials such as formability, perovskite stability, bandgap, and Power Conversion Efficiency (PCE).[11–13] In this context, Variational Autoencoders (VAEs) have received a lot

* Corresponding author
‡These authors contributed equally to this work
† Electronic Supplementary Information (ESI) available: See DOI: 00.0000/00000000.
[a]*Chemistry Department, Johns Hopkins University, Baltimore, MD, USA; E-mail: r.hernandez@jhu.edu*
[b]*Chemical and Biomolecular Engineering Department, Johns Hopkins University, Baltimore, MD, USA*
[c]*Materials Science and Engineering Department, Johns Hopkins University, Baltimore, MD, USA*

of attention in the past couple of years due to their ability to construct latent spaces defined by prior probability distributions which are often supervised for the material or molecule property targeted for prediction.[14] Equipped with an appropriate sampling strategy, one can begin traversing and sampling points from regions of the latent space, and then decode them back to the original feature representation space. VAEs have been showcased in designing molecules,[15] Perovskite Solar Cell (PSC) bandgap prediction[16,17] and PSC PCE prediction.[18]

In this work, we develop an alternate strategy for designing materials—and MHPs in particular—using a Nested Autoencoder (NestedAE) framework in which we connect a a series of Autoencoders (AEs) representing different scales. The successive transfer of information from one AE to the next has previously been used to build deep neural networks by transferring information from one AE to the next.[19] In this pretraining strategy, deep neural nets are built layer-wise in such a way that each added hidden layer is trained to reconstruct the inputs coming from the layer before it, and thus essentially acts as an AE. By initializing weights of the hidden layers closer to the local optimum, this greedy layer-wise unsupervised training was found to yield better accuracy at classification tasks similar to the ones we are targeting.[19–21] Another way to improve the training of deep neural networks for such a classification task is to jointly train each hidden layer for both classification and reconstruction. By reducing the dimensionality of the hidden layers, one can arrive at compact representations of the inputs.[22]

The NestedAE method is distinct from the previous approaches noted above in two primary ways: First, in contrast to a single deep neural network built on an entire database, each AE in the NestedAEs has a distinct architecture and is trained on its own dataset. Second, in previous approaches, the decoder only has to reconstruct the input coming from the previous hidden layer while in NestedAE the AE has to reconstruct both the latents from the previous (or finer-grained) AE in addition to the features from the dataset it is trained on. In Sec. 3.1, we demonstrate the efficacy of this hierarchical approach using a synthetic database. Specifically, we benchmark the use of NestedAE in terms of reconstruction errors and robustness to noisy data.

We then demonstrate the applicability of NestedAE to the design of MHPs. These materials are a class of perovskite crystals where the A site is occupied by an organic or inorganic cation; *e.g.*, cesium (Cs), methylammonium (MA) or formamidinium (FA). The B site is occupied by a transition metal cation usually Sn or Pb. The X site is occupied by a halide usually Cl, Br or I. The rise in popularity of these crystals as solar cell materials is fueled by their ease of synthesis and promising photovoltaic properties. The highest reported efficiency for a 1 cm^2 single-junction perovskite solar cell is 24.35 % while that for a 1 cm^2 perovskite/Si multi junction solar cell is 33.7 %.[23] Their high specific power (*viz.* power to weight ratio), compatibility with flexible substrates and radiation resistance properties have led to recent interest for their use in space technologies.[24] They can also be synthesized to take on a variety of morphologies. 0D perovskites or perovskite quantum dots are isolated metal halide octahedra surrounded by organic or inorganic cations. 1D perovskites or perovskite nanowires are metal halide octahedra connected in a chain. 2D perovskites have sheets of metal halide octahedra spaced by large organic cations. 3D perovskites which are the focus of this study form a 3D network of metal halide octahedra with inorganic cations or small organic cations occupying the voids of the network. Taken together, this leads to an overwhelming number of material compositions and processing conditions in optimizing the selection of a suitable MHP as is seen in most material design problems. Intuition may guide the choice toward a lower-dimensional subset of compositions and processing conditions favorable for fabricating a device with high performance. However, it is not *a priori* obvious how varying these choices will affect the performance of the device, or how to unravel what is likely to be a highly nonlinear correlation between them. For example, properties of the ions in the MHP crystal affect charge transport properties in the 3D MHP metal halide network,[25], and the type of extracting layers used in the device stack affects how efficiently these charges can be extracted to their respective contacts. How these properties at different physical length scales synergize to affect device performance remains unknown. In Sec. 3.2, we show how the 2D latent space of the last AE in NestedAE can be used to design a PSC for a given PCE.

Thus, the NestedAE method creates a bridge across material properties measured at multiple physical scales using data driven techniques. It is an interpretable machine learning model that uses AE architectures as building blocks. Each AE represents a single physical scale, and information is transferred between pairs of scales via the latent space of the finer-grained scale AE This approach is somewhat analogous to value iteration in a Markov Decision Process (MDP)[26] in the sense that information is transferred as subsequent AEs depend on the latents of the preceding AE. More generally, the NestedAE framework also includes physical properties of the current scale among its inputs.

In summary, below, we first develop the use of the theory for NestedAE (Sec. 2.1) on a synthetic multiscale dataset (Sec. 2.2), and on a multiscale MHP dataset constructed from multiple sources (Sec. 2.3). The ensuing results for these two datasets in Secs. 3.1 and 3.2, respectively suggest that the NestedAE method presented here can be applied to any type of multiscale materials with distinct properties at different physical scales.

2 NestedAE

2.1 Theory and Formulation

NestedAE is a novel neural network architecture designed to build multi-scale material models by bridging material properties measured/computed at different physical scales. The building blocks for the NestedAE method are the AEs at each of the scales. For example, let us consider N physical length scales. This implies we have N sequentially trained AEs. At each scale, the encoder-decoder is trained and the latent space is learned. For the subsequent scale, the input features include the latent variables from the previous scale and additional physical properties from the current scale. This process of *"nesting*" latents within the latents of the successive autoencoder can be continued for the N autoencoders. Figure 1 gives an overview of this architecture when

trained in an unsupervised and supervised manner. Details regarding the training strategies and loss functions used for training on the synthetic dataset are described in Sec. 2.2.

For the MHP dataset, the NestedAE architecture involves supervising the latent space for prediction of MHP properties. In this architecture, a predictor, which is a feed forward neural network is attached to the latent space of the autoencoder at a given scale. The inputs to the predictor are the latent variables and the output is the MHP property, for example, the band gap. In this training scheme, the encoder, decoder and predictor are trained simultaneously in a multi-task learning format as described in Sec. 2.3 and Algorithm 2.

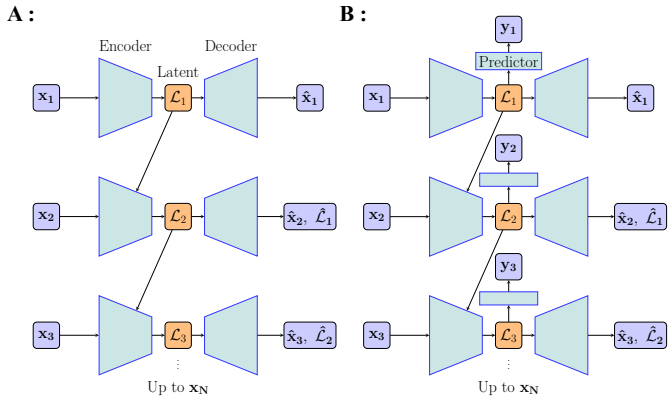

Fig. 1 Unsupervised (A) and supervised (B) NestedAE architecture.

2.2 NestedAE architecture and training scheme for synthetic dataset

In an unsupervised learning setting, the training of a NestedAE is the natural extension of that for a traditional AE. Each of the AEs in the NestedAE represents a particular scale, and they are trained sequentially. The training scheme is demonstrated on the synthetic dataset and the results are reported in Sec. 3.1. In this illustrative example, the first—viz finest-grained or lowest scale—AE consists of two hidden layers with 4 and 3 nodes, respectively. The latent space dimension is fixed to 2. Tangent hyperbolic activation functions are employed in the hidden layers. In order to prevent overfitting of the model, the weights are explored through L1 regularization. The second (scale) AE consists of three sequential hidden layers containing 9,7 and 6 nodes, respectively, and uses the same activation function and regularization scheme as the first AE. The loss functions and training scheme are shown in Algorithm 1.

Algorithm 1 Unsupervised learning scheme for NestedAE

Require: Multi-scale dataset

1: Initialize weights : $\mathbf{W^1_{encoder}}$, $\mathbf{W^1_{decoder}} \leftarrow$ xavier_normal()

2: Optimize encoder and decoder

$J = \frac{1}{10}\sum_{b=1}^{10}|\mathbf{X^1_{prop}} - \hat{\mathbf{X}}^{\mathbf{1}}_{\mathbf{prop}}| + 0.01\|\mathbf{W^1_{encoder}}\|_1$

2.3 NestedAE architecture and training scheme for multi-scale MHP

The first and second stage AEs in NestedAE are very similar in design in that they have single hidden layers for the encoder and decoder, and use a hyperbolic tangent activation for the encoder to create a bounded latent space with values along each dimension ranging from 1 to -1. L1 regularization is employed on the encoder weights to introduce sparsity in feature selection. A regularization coefficient of 0.01 is used. Bandgap and PCE are predicted from the latents of the first and second stage AEs, respectively. The prediction neural networks are comprised of 2 hidden layers with ReLU activation.[27] After the first hidden layer, we use dropout regularization with a dropout probability of 0.1.[28] This implementation prevents overfitting in the bandgap and PCE prediction. A multi task optimization scheme is employed in which the encoder, decoder and predictor neural networks are trained together. The prediction and reconstruction losses in both AEs are measured through Mean Absolute Error (MAE) loss functions. The mini-batch gradient descent algorithm is used during training with batch sizes of 10 and 100 for the first and second AE, respectively, and the Adam optimizer with a learning rate of 0.001 in both. Training was stopped in both AEs after reaching 1500 epochs as we observed no significant reduction in validation loss thereafter. All neural network architectures are trained using the PyTorch Lightning package. The numerical performance and applications of the NestedAE method are reported below.

Algorithm 2 NestedAE training scheme for multiscale MHP dataset

Require: Density Functional Theory (DFT) bandgap dataset
Features : A,B,X site ionic properties ($\mathbf{X^1_{prop}}$)
Labels : DFT predicted bandgap ($\mathbf{y^1_{bandgap}}$)

Require: PSC device dataset
Features : A,B,X site ionic properties ($\mathbf{X^2_{prop}}$), Electron Transporting Layer (ETL) ($\mathbf{X^2_{ETL}}$) and Hole Transporting Layer (HTL) ($\mathbf{X^2_{HTL}}$) layer types
Labels : PCE ($\mathbf{y^2_{PCE}}$)

1: Initialize weights : $\mathbf{W^1_{encoder}}$, $\mathbf{W^1_{decoder}}$, $\mathbf{W^1_{predictor}} \leftarrow$ xavier_normal()

2: Optimize encoder, decoder and predictor weights of autoencoder 1 for predicting DFT bandgaps by training on the DFT bandgap dataset. Use mini-batch gradient descent with the following multi-task objective function:
$J = \frac{1}{10}\sum_{b=1}^{10}|\mathbf{X^1_{prop}} - \hat{\mathbf{X}}^{\mathbf{1}}_{\mathbf{prop}}| + \frac{1}{10}\sum_{b=1}^{10}|\mathbf{y^1_{bandgap}} - \hat{\mathbf{y}}^{\mathbf{1}}_{\mathbf{bandgap}}| + 0.01\|\mathbf{W^1_{encoder}}\|_1$

3: Predict DFT bandgaps and latents (L^1) for the MHP compositions in PSC device dataset.

4: Initialize weights : $\mathbf{W^2_{encoder}}$, $\mathbf{W^2_{decoder}}$, $\mathbf{W^2_{predictor}} \leftarrow$ xavier_normal()

5: Optimize encoder, decoder and predictor weights of autoencoder 2 for predicting PCE by training on the PSC device dataset and latents from AE 1 (L^1). Use mini-batch gradient descent with the following multi-task objective function:
$J = \frac{1}{100}\sum_{b=1}^{100}|\mathbf{L^1} - \hat{\mathbf{L}}^{\mathbf{1}}| + \frac{1}{100}\sum_{b=1}^{100}|\mathbf{y^2_{PCE}} - \hat{\mathbf{y}}^{\mathbf{2}}_{\mathbf{PCE}}| +$ Cross_Entropy($\mathbf{X^2_{ETL}}, \hat{\mathbf{X}}^{\mathbf{2}}_{\mathbf{ETL}}$) + Cross_Entropy($\mathbf{X^2_{HTL}}, \hat{\mathbf{X}}^{\mathbf{2}}_{\mathbf{HTL}}$) + $0.01\|\mathbf{W^2_{encoder}}\|_1$

3 Results

3.1 NestedAE performance on Synthetic dataset

The performance of the NestedAE architecture has been benchmarked using multi-dimensional datasets composed of analytical functions. We refer to these datasets as the synthetic datasets in the remainder of the paper. The first synthetic dataset is composed of 13 functions ($f_i \quad i = 1,2,\ldots 13$) of 5 independent variables, $z_i \quad i = 1,2,\ldots 5$, as available in Sec. S1 of the Supplementary Material (SM). Four functions depend only on z_1 and z_2, and these form the input feature space for the first AE in the NestedAE architecture. The remaining seven functions depend on all five independent variables and form the input feature space for the second AE in the NestedAE architecture. The second dataset consists of the same 13 functions with Gaussian noise added to each. The aim of the second dataset is to assess the robustness of the NestedAE architecture to noisy inputs such as would be expected from an experimental data set. The application of NestedAE on these synthetic datasets represents a top-down approach in the sense of coarse grain modeling since the inputs to the first AE are functions of fewer variables (z_1 and z_2) while the inputs of the second AE are functions to all independent variables. This is contrary to the bottom-up approach used in describing the logic of the NestedAE construction in the sensed that the first scale had therein been described as being the fine-grained scale. However, the benchmarking performed on the synthetic database remains relevant because of the symmetry between the two frameworks. Further details for all entries in the synthetic dataset are given in Sec. S1 in the SM.

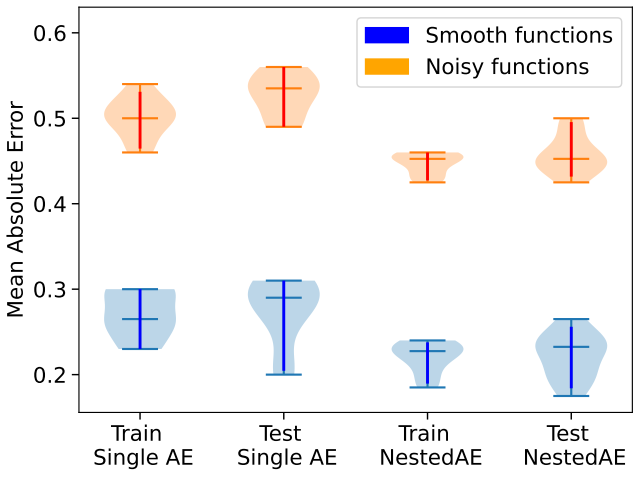

Fig. 2 Comparison of MAE between NestedAE and single AE architectures for the training and test datasets. The training is done using ten different random seeds. The blue violins show the reconstruction error distributions for 13 analytical functions created as a function of 5 independent variables, z_i, $i = 1,2,\ldots 5$. The orange violins show the reconstruction error distributions for the 13 analytical functions with additive Gaussian noise.

Figure 2 compares the performance of a single AE and NestedAE for reconstruction errors, and robustness to noise. These results are generated by training the AE architectures from 10 different random seeds used to initialize variations in the test-train data split and initialization of the AE weights. The data with lower reconstruction errors shown using blue violins correspond to the dataset with analytical functions and the data with higher reconstruction errors shown using orange violins correspond to the dataset generated from noisy functions. In both datasets, we see that the NestedAE approach outperforms the single AE architecture. This result is primarily due to the splitting up of the input vector into two input feature vectors across the two AEs in NestedAE and the resulting reduction of the dimensions of each input vector into the latent space between the sequential AEs. In doing so, we allow more information to be encoded within a (latent) space that has the same effective dimensionality, thereby reducing the reconstruction errors. We also see that the NestedAE approach is more robust to noise and results in overall smaller reconstruction errors with better convergence. The correlations between the principal components of the latent spaces obtained from NestedAE and single AE are also available in Figs. S1, S2, and S3 in the SM. It is difficult to confirm the relative dimensionality between these spaces. However, the latent space of the first stage of the NestedAE appears to have a significant contraction to a reduced dimensional curve indicating the corresponding reduced dimensionality of the underlying domain space used to generate the synthetic database.

3.2 NestedAE performance on multiscale Metal Halide Perovskites dataset

In describing material properties, we may ultimately need to include many different scales explicitly, and hence a NestedAE framework would include several corresponding AEs. Each of these AEs would then be trained sequentially, allowing the machine to perform the coarse-graining of information at each scale resulting in the corresponding latent space for each AE. For simplicity, here we focus on a NestedAE trained on data at only two different material length scales. The first-stage AE is trained for bandgap prediction. For this, we use the dataset published by Mannodi-Kanakkithodi et al.[29] The dataset consists of properties of ions situated at A, B and X sites: ionic radius, weight, electronegativity, ionization energy electron affinity, and others. This dataset includes 499 pure and mixed MHPs with mixed compositions at all three sites. For each composition, Mannodi-Kanakkithodi et al.[29] computed the bandgap using a Perdew-Burke-Ernzerhof (PBE) functional and Projector-Augmented Wave (PAW) pseudo potential. They found surprisingly good accuracy using this level of theory, but we recognize that the accuracy of the machine learning approach introduced here is subject to that of our selected database. That is, improvements on the present work could arise from the introduction of databases generated by experiment or through more accurate levels of theory.

The second stage AE is trained for PCE prediction. For this, we use the perovskite device dataset published by Jacobsson et al.[30] The original dataset consists of devices manufactured with different perovskite formulations and processing conditions. Each device is characterized for its current density-voltage (J-V) performance, various device stabilities and outdoor testing. We chose to

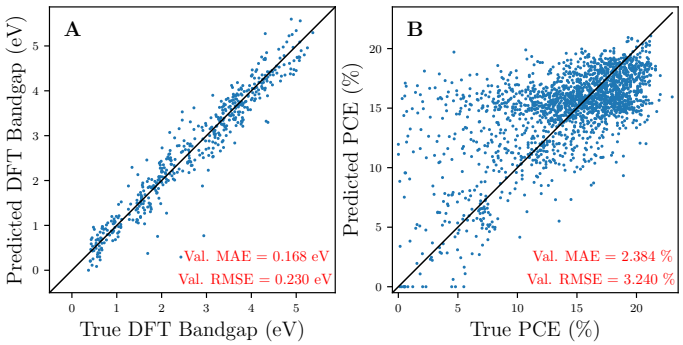

Fig. 3 Plots depicting NestedAE predictions for (A) DFT bandgap and (B) PCE versus true DFT bandgap and PCE. Model predictions are from the test dataset.

focus only on the effect of the types of contacts on the PCE. In particular, we chose the PCE measured in the reverse scan setup as a majority of the devices in the dataset were characterized accordingly. We filtered for samples from the original dataset that were manufactured using a single-step spin coating process followed by gas quench-induced crystallization. This type of quenching was found to produce uniform films with densely packed grains circumventing the formation of pinholes.[31] J-V performance is a standard way to measure PCE of solar cell devices, hence we also filtered for devices that did not meet the standard protocol for measuring J-V performance. The standard protocol includes measurements performed at 25°C, under ambient or dry air conditions, using light intensity of 100 mW/cm2, and employing light not passed through a mask. We dropped duplicates and devices from the dataset which did not report a PCE measurement. We also dropped devices which involved particular ETL or HTL layers that are seen in fewer than 5% of the devices in the dataset. This was done to limit the size of the one hot encoding to only the most popular types of ETL and HTL layers in the reported materials. After applying all these filters, we were left with a dataset containing 2018 devices. The dataset contains 7 types of ETL (C_{60}/Bathocuproine (BCP), [6,6]-Phenyl-C61-butyric acid methyl ester (PCBM-60), PCBM-60/BCP, SnO_2-compact(c), SnO_2-nanoparticle(np), TiO_2-c, TiO_2-c/TiO_2-mesopororus(mp)) and 4 types of HTL (NiO-c, PEDOT:PSS, PTAA, Spiro-MeOTAD).

In optimizing the NestedAE architecture for this multiscale dataset, we found that a latent space dimensionality of at least 12 is required for autoencoder 1. Any reduction in the dimension beyond that resulted in an increase in both reconstruction and prediction losses. Five-fold cross-validation was implemented to assess the model performance for overfitting. The average train and validation MAE across the 5 folds in bandgap prediction is shown in Table 1. This is on par with bandgap prediction accuracy observed by Mannodi-Kanakkithodi et al.[29] The performance of supervised AE performance was also found by Lei et al.[32] to be on par with standard neural network architectures when comparing prediction metrics. Figure 3 shows that most of the bandgaps are predicted close to their true values.

Supervised latents from the first stage AE are passed to the encoder of the second stage AE. Among the five models trained using the 5 folds from five-fold cross-validation, the model which has the smallest sum of the validation bandgap MAE and property reconstruction MAE is chosen to predict the bandgaps and latents for the compositions in the perovskite device dataset. In training the second AE, we found that a latent space dimensionality of at least 18 is required. The average train and validation MAE in PCE prediction is shown in 1. We trained a single AE using only the type of ETL and HTL to predict PCE. We observed the average train and validation MAE for PCE prediction to be higher when compared to using the NestedAE approach. From this, we observe that the information encoded in the supervised latents improves the PCE prediction. It should be noted however that the ETL type predictions are slightly better in single AE than compared to NestedAE.

Single AE does not provide information about bandgaps because the perovskite device dataset does not contain bandgap information for all formulations. Approximately 25% of compositions have missing bandgap values. Even when there is available bandgap information, the various methods used to measure the bandgap are sometimes inconsistent or insufficiently detailed. Hence, the exact same device can be associated with multiple bandgap values which impedes training of a single AE. Figure 3 shows that there is a large spread in PCE predictions at low PCE values but the NestedAE performs better at predicting high PCE values. This could be attributed to a smaller number of low PCE data points in the perovskite device dataset for the NestedAE model to learn from. Li et al[33] have also created a model for PCE prediction using bandgap, the energy difference between HOMO of HTL and HOMO of perovskite materials, and the energy difference between LUMO of ETL and LUMO of perovskite materials as features. They were able to achieve an RMSE in PCE prediction (in % units) of 3.27 % using the predicted bandgap. While this is on par with the PCE RMSE values predicted from NestedAE, we are also able to predict the ETL and HTL types that provide such PCE values.

Figure 5 shows the 2D latent space for NestedAE. Figure 5B shows the 2D latent points for each perovskite device colored according to the NestedAE predicted PCE value. We see a clear distinction between points having low PCE (lying near the bottom left corner of the plot) and points having high PCE (lying near the top left corner). Figure 5A shows the 2D latent space colored according to the values of the DFT bandgap predictions. Most of the devices in the dataset were constructed using $MAPbI_3$ and hence the majority of the NestedAE predicted DFT bandgaps are around 2.55 eV. Figures 5C and 5D show the 2D latent space colored according to the predicted ETL and HTL type. Here, we see a clear segregation between organic and inorganic transport layers. This is in contrast to Figs. 4 B and D. Therein, we plot the bandgaps vs PCE for devices in the perovskite dataset and color the points according to the type of ETL and HTL layer used respectively. It is not obvious from the PCE and bandgap values in the naive figures how the ETL and HTL layers should be clustered to produce better devices. In contrast, the resulting latent spaces in the NestedAE do provide the necessary segregation to infer such design.

The findings from the 2D latent space of the NestedAE,—and

	NestedAE		Single AE (No A,B,X prop.)	
Property	**Train**	**Validation**	**Train**	**Validation**
E_{gap}^{PBE} (eV)	0.156 ± 0.007	0.168 ± 0.027	NA	NA
A,B,X	0.021 ± 0.001	0.019 ± 0.002	NA	NA
PCE (%)	2.566 ± 0.026	2.384 ± 0.034	3.215 ± 0.041	2.889 ± 0.118
Latent	0.032 ± 0.005	0.029 ± 0.01	NA	NA
ETL type	82.341 ± 0.093	81.846 ± 0.105	95.122 ± 0.017	96.145 ± 0.011
HTL type	99.525 ± 0.002	99.61 ± 0.003	100 ± 0	100 ± 0

Table 1 Comparison of prediction metrics for properties using NestedAE compared to single AE. MAE is the metric used for measuring E_{gap}^{PBE}, A, B, X ionic properties, PCE and latent prediction errors. ETL type and HTL type are categorical variables hence we use accuracy to report their prediction metrics. NestedAE was trained using 5-fold cross-validation. We report the mean and standard deviations for each prediction across the 5 folds.

from Figs 5 B, C and D, in particular—corrobate findings from the MHP literature. In general, organic ETL layers (C60, PCBM) hinder PCE while inorganic ETL layers (TiOx, SnOx) have been shown to provide improved efficiencies due to higher stability and easily adjustable electronic and optical properties.[34] This is reflected in Figs 5 B and C where blue, orange and green dots (which correspond to organic ETL layers) in Fig 5 C occupy the same region in Fig 5 B corresponding to a region of low PCE. Similarly the red, purple, brown and grey dots (which correspond to inorganic ETL layers) in Fig 5 C occupy the high PCE region in Fig 5 B. Among HTL layers, NiO has low conductivity causing hole accumulation and recombination at perovskite/NiO layer.[35]. Spiro-MeOTAD and PTAA on the other hand possess good hole conductivity and have good band alignment with the valance bands of a majority of the perovskite layers.[36] This is reflected in Figs 5 B and D where the yellow and cyan dots (which correspond to NiO and PEDOT:PSS HTL layers) in Fig 5 D occupy the low PCE region in Fig 5 B. Similarly the magenta and dark green dots (which correspond to PTAA and Spiro-MeOTAD HTL layers) in Fig 5 D occupy the high PCE region in Fig 5 B. Using a combination of inorganic ETL layer, Spiro-MeOTAD as a HTL layer, and low bandgap formulations, we thus lie in a goldilocks zone of high PCE perovskite solar cells.

4 Conclusion

In physics-based models representing multiple scales, there is a transfer of information from one model to the next through a parameter or set of parameters measured from the lower-scale model. In NestedAE, our models are AEs which construct a latent representation of the original feature space and the latent information is passed onto successive AEs that represent larger physical scales. In an illustrative example of this work, the first AE learned how to predict DFT bandgaps using the atomic and ionic properties of the A, B and X ions in perovskite crystals. These are used to infer DFT bandgap values of compositions in the perovskite device dataset.[30] In turn, they help to complete the missing values in the dataset and to predict the PCE values through the supervised latents. Possible aims for future research using NestedAE include characterization of the impact from adding more supervised AEs, prediction of film and charge transport properties, and determination of how they affect PCE. Notably, the use of NestedAEs is not limited to tabular data, and this accommodates its use in learning from complex data sets.

One of the most striking features of the proposed architecture is that it can be used in completing a database of materials which have missing entries. For instance in the case of perovskites, let us consider that we have data for $MAPbI_3$ for AE1 and AE2, and that both AEs are trained sequentially for this perovskite material. Suppose further that we have another perovskite material ($FAPbBr_3$) for which we have the input feature set for AE1 (i.e., x_1) but not for AE2 (i.e., x_2). In this case, we can obtain our latent $\mathscr{L}_1$ for AE1 by feeding in x_1. However, since x_2 is not known, we the sample $\mathscr{L}_2$ that we have trained using known data samples and explore the space such that the decoder's $\hat{\mathscr{L}}_1$ is approximately equal to $\mathscr{L}_1$ for $FAPbBr_3$ within an error tolerance. As such, NestedAE offers a viable model enabling multiscale materials discovery through a search of the associated reduced-dimensional latent spaces rather than the full-dimensional feature space.

5 Data and code availability

Datasets and code for NestedAE is available on GitHub at https://github.com /rxhernandez/NestedAE.

Author Contributions

Nikhil Thota : Downstream Conceptualization, Methodology, Data Curation, Investigation, Software, Validation, Visualization, Writing - Original Draft, Writing - Review & Editing, **Dr. Maitreyee Sharma Priyadarshini :** Downstream Conceptualization, Formal Analysis, Methodology, Validation, Visualization, Writing - Original Draft, Writing - Review & Editing, **Dr. Rigoberto Hernandez :** Initial and Downstream Conceptualization, Methodology, Formal Analysis, Visualization, Writing - Review and Editing, Supervision, and Funding Acquisition.

Conflicts of interest

There are no conflicts to declare.

6 Acknowledgments

This work has been primarily supported by the U. S. Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES), under Award #DE-SC0022305 (formulation engineering of energy materials via multiscale learning spirals). The computing resources necessary for this work were performed in part on Bridges-2-GPU at the Pittsburgh Supercomputing Center through allocation CTS090079 provided by Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS), which is supported by National Science Foundation (NSF) grants

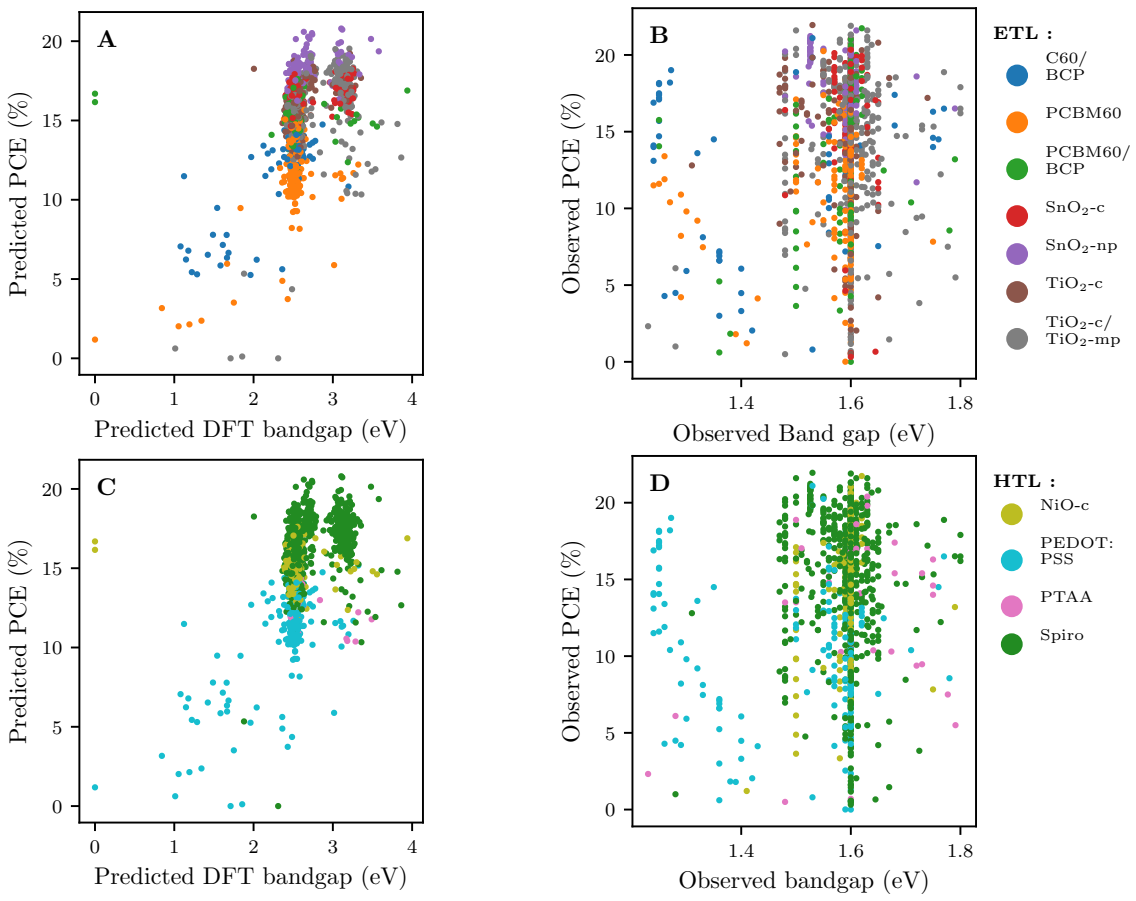

Fig. 4 Predicted (A and C) and observed (B and D) PCE vs DFT bandgap. The points are coded according to ETL types (A and B) as listed in the legend at the top right, and HTL types (C and D) as listed in the legend at the bottom right.

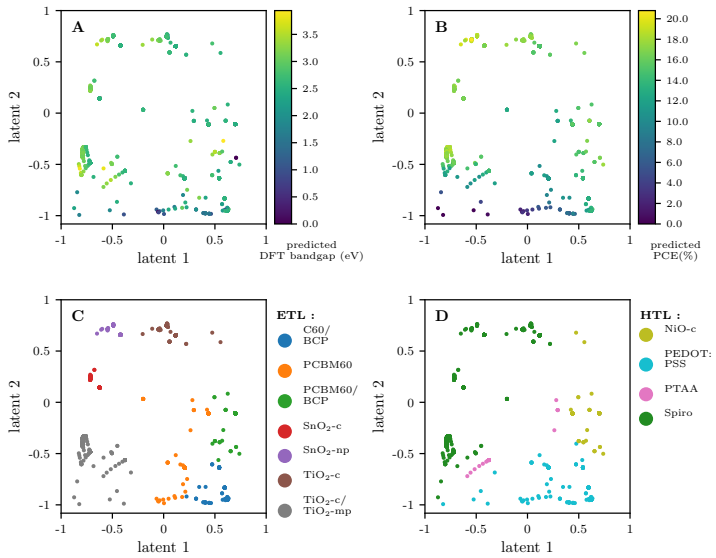

Fig. 5 Two-dimensional latent space of the NestedAE color-coded according to the predicted DFT bandgap (A), the predicted PCE (B), the corresponding ETL type (C) as listed in the legend, and the corresponding HTL type (D) as listed in the legend. The latent space is shown for the model that achieved highest accuracy in classifying the ETL and HTL material type among the 5 models trained through 5 fold cross validation on the perovskite device dataset.

#2138259, #2138286, #2138307, #2137603, and #2138296. Additional computing resources were provided by the Advanced Research Computing at Hopkins (ARCH) high-performance computing (HPC) facilities supported by the NSF MRI Grant (OAC-1920103).

References

1 I. R. E. Agency, *World Energy Transitions Outlook 2023: 1.5°C Pathway*, 2023, https://www.irena.org/Publications/2023/Jun/World-Energy-Transitions-Outlook-2023, Accessed on September 13, 2023.

2 G. C. Peng, M. Alber, A. B. Tepole, W. R. Cannon, S. De, S. Dura-Bernal, K. Garikipati, G. Karniadakis, W. W. Lytton, P. Perdikaris, L. Petzold and E. Kuhl, *Arch. Comput. Methods Eng.*, 2021, **28**, 1017–1037.

3 A. E. Clark, H. Adams, R. Hernandez, A. I. Krylov, A. M. N. Niklasson, S. Sarupria, Y. Wang, S. M. Wild and Q. Yang, *ACS Cent. Sci.*, 2021, **7**, 1271.

4 J. Fish, G. J. Wagner and S. Keten, *Nat. Mater.*, 2021, **20**, 774–786.

5 S. Schmauder and I. Schäfer, *Mater. Today*, 2016, **19**, 130–131.

6 W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das and H. C. Andersen, *J. Chem. Phys.*, 2008, **128**, 244114.

7 F. Cipcigan, J. Crain, V. Sokhan and G. J. Martyna, *Rev. Mod. Phys.*, 2019, **91**, 025003.

8 J. Jin, A. J. Pak, A. E. Durumeric, T. D. Loose and G. A. Voth, *J. Chem. Theory Comput.*, 2022, **18**, 5759–5791.

9 S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, *J. Phys. Chem. B*, 2007, **11**, 7812–7824.

10 X. Periole, M. Cavalli, S.-J. Marrink and M. A. Ceruso, *J. Chem. Theory Comput.*, 2009, **5**, 2531–2543.

11 Q. Tao, P. Xu, M. Li and W. Lu, *npj Comput. Mater.*, 2021, **7**, 23.

12 Z. Hui, M. Wang, X. Yin, Y. Wang and Y. Yue, *Comput. Mater. Sci.*, 2023, **226**, 112215.

13 Y. Xu, J. Ge and C.-W. Ju, *Energy Adv.*, 2023.

14 D. P. Kingma and M. Welling, *arXiv preprint arXiv:1312.6114*, 2013.

15 R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams and A. Aspuru-Guzik, *ACS Cent. Sci.*, 2018, **4**, 268–276.

16 H. Choubisa, M. Askerka, K. Ryczko, O. Voznyy, K. Mills, I. Tamblyn and E. H. Sargent, *Matter*, 2020, **3**, 433–448.

17 C. J. Court, B. Yildirim, A. Jain and J. M. Cole, *J. Chem. Inf. Model.*, 2020, **60**, 4518–4535.

18 Z. Ren, F. Oviedo, H. Xue, M. Thway, K. Zhang, N. Li, J. D. P. , M. Layurova, Y. Wang, S. Tian, T. Heumueller, E. Birgersson, F. Lin, A. Aberle, S. Sun, I. M. Peters, R. Stangl, C. J. Brabec and T. Buonassisi, 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC), 2019, pp. 3054–3058.

19 Y. Bengio, P. Lamblin, D. Popovici and H. Larochelle, Adv. Neural Inf. Process Syst., 2006.

20 P. Vincent, H. Larochelle, Y. Bengio and P.-A. Manzagol, Proceedings of the 25th International Conference on Machine Learning, 2008, pp. 1096–1103.

21 D. Erhan, A. Courville, Y. Bengio and P. Vincent, Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010, pp. 201–208.

22 M. Ranzato and M. Szummer, Proceedings of the 25th international conference on Machine learning, 2008, pp. 792–799.

23 M. A. Green, E. D. Dunlop, M. Yoshita, N. Kopidakis, K. Bothe, G. Siefer and X. Hao, *Prog. Photovolt. Res. Appl.*, 2023, **31**, 651–663.

24 Y. Tu, J. Wu, G. Xu, X. Yang, R. Cai, Q. Gong, R. Zhu and W. Huang, *Adv. Mater.*, 2021, **33**, 2006545.

25 Y.-C. Hsiao, T. Wu, M. Li, Q. Liu, W. Qin and B. Hu, *J. Mater. Chem. A*, 2015, **3**, 15372–15385.

26 R. Bellman, *Journal of mathematics and mechanics*, 1957, 679–684.

27 X. Glorot, A. Bordes and Y. Bengio, Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011, pp. 315–323.

28 G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. R. Salakhutdinov, *arXiv preprint arXiv:1207.0580*, 2012.

29 A. Mannodi-Kanakkithodi and M. K. Y. Chan, *Energy Environ. Sci.*, 2022, **15**, 1930–1949.

30 T. J. Jacobsson, A. Hultqvist, A. García-Fernández, A. Anand, A. Al-Ashouri, A. Hagfeldt, A. Crovetto, A. Abate, A. G. Ricciardulli, A. Vijayan, A. Kulkarni, A. Y. Anderson, B. P. Darwich, B. Yang, B. L. Coles, C. A. R. Perini, C. Rehermann, D. Ramirez, D. Fairen-Jimenez, D. D. Girolamo, D. Jia, E. Avila, E. J. Juarez-Perez, F. Baumann, F. Mathies, G. S. A. González, G. Boschloo, G. Nasti, G. Paramasivam, G. Martínez-Denegri, H. Näsström, H. Michaels, H. Köbler, H. Wu, I. Benesperi, M. I. Dar, I. Bayrak Pehlivan, I. E. Gould, J. N. Vagott, J. Dagar, J. Kettle, J. Yang, J. Li, J. A. Smith, J. Pascual, J. J. Jerónimo-Rendón, J. F. Montoya, J.-P. Correa-Baena, J. Qiu, J. Wang, K. Sveinbjörnsson, K. Hirselandt, K. Dey, K. Frohna, L. Mathies, L. A. Castriotta, M. H. Aldamasy, M. Vasquez-Montoya, M. A. Ruiz-Preciado, M. A. Flatken, M. V. Khenkin, M. Grischek, M. Kedia, M. Saliba, M. Anaya, M. Veldhoen, N. Arora, O. Shargaieva, O. Maus, O. S. Game, O. Yudilevich, P. Fassl, Q. Zhou, R. Betancur, R. Munir, R. Patidar, S. D. Stranks, S. Alam, S. Kar, T. Unold, T. Abzieher, T. Edvinsson, T. W. David, U. W. Paetzold, W. Zia, W. Fu, W. Zuo, V. R. F. Schröder, W. Tress, X. Zhang, Y.-H. Chiang, Z. Iqbal, Z. Xie and E. Unger, *Nat. Energy*, 2022, **7**, 107–115.

31 A. Babayigit, J. D'Haen, H.-G. Boyen and B. Conings, *Joule*, 2018, **2**, 1205–1209.

32 L. Le, A. Patterson and M. White, *Adv. Neural Inf. Process Syst.*, 2018, **31**, 1–11.

33 J. Li, B. Pradhan, S. Gaur and J. Thomas, *Adv. Energy Mater.*, 2019, **9**, 1901891.

34 B. Yang, S. Peng and W. C. H. Choy, *EcoMat*, 2021, **3**, e12127.

35 X. Zhao, W. Zhang, X. Feng, X. Guo, C. Lu, X. Li and J. Fang, *Chemical Engineering Journal*, 2022, **435**, 135140.

36 F. M. Rombach, S. A. Haque and T. J. Macdonald, *Energy Environ. Sci.*, 2021, **14**, 5161–5190.