Predicting AFM Topography From Optical Microscopes Using Deep-Learning

Jaewoo Jeong[1], Taeyeong Kim[1], Bong Jae Lee[1], and Jungchul Lee[1]

[1]Affiliation not available

September 19, 2022

Abstract

Atomic force microscopy (AFM) is routinely used as a metrological tool among diverse scientific and engineering disciplines. A typical AFM, however, is intrinsically limited by low throughput and is inoperable under extreme conditions. Thus, this work attempts to provide an alternative with a conventional optical microscope (OM) by training a deep learning model to predict surface topography from surface OM images. The feasibility of our novel methodology is shown with germanium-on-nothing (GON) samples, which are self-assembled structures that undergo surface and sub-surface morphological transformations upon high-temperature annealing. Their transformed surface topographies are predicted based on OM-AFM correlation of 3 different surfaces, bearing an error of about 15% with 1.72× resolution upscale from OM to AFM. The OM-based approach brings about significant improvement in topography measurement throughput (equivalent to OM acquisition rate, up to 200 frames per second) and area ($\sim$1 mm^2). Furthermore, this method is operable even under extreme environments when an *in-situ* measurement is impossible. Based on such competence, we also demonstrate the model's simultaneous application in further specimen analysis, namely surface morphological classification and simulation of dynamic surfaces' transformation.
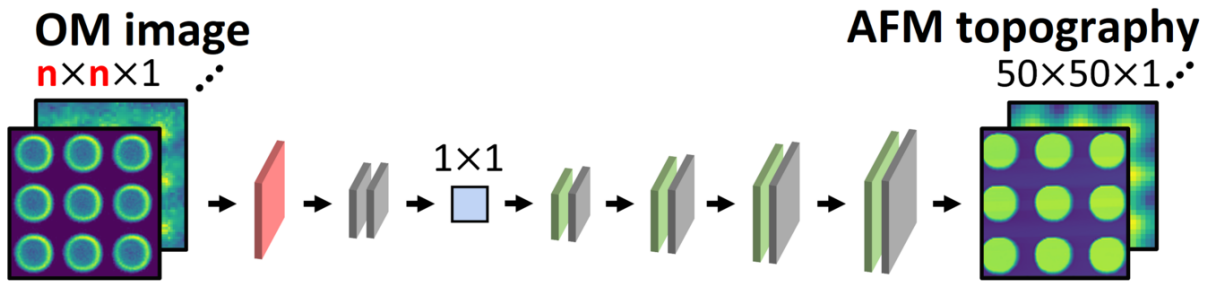
[a] Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology
291, Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea
[b] Center for Extreme Thermal Physics and Manufacturing, Korea Advanced Institute of Science and Technology
291, Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea
Email Address: bongjae.lee@kaist.ac.kr, jungchullee@kaist.ac.kr

Introduction

Since its inception, Atomic Force Microscopy (AFM) has been ubiquitously used for probing nanoscale surfaces in various domains: biology (Garcia et al., 2020; Ke et al., 2008), material science (Labidi et al., 2017), chemistry (Fan & Chelikowsky, 2021), and physics (Xue et al., 2021), to name a few. Indeed, its measurement capability in the sub-optical wavelength regime has been paramount to establishing a deeper understanding in each scientific domain. Outside of academia, AFM has also been exploited in semiconductor industries for surface metrology and inspection (Sadeghian et al., 2017; Deng et al., 2021). Today, along with scanning electron microscopy (SEM) and transmission electron microscopy (TEM), AFM continues to remain as a prevalent tool for scrutinizing surface topography in a sub-nanometer scale for diverse needs in academic and industrial domains.

Despite its invaluable advantages in measurement resolution, the raster scanning approach of AFM intrinsically limits the measurement throughput compared to optical measurements. As a solution, new raster scanning methodologies have been proposed to achieve a video frame AFM in the range of 20 to 45 frames per second (FPS). However, the significance of the advantages achieved by these approaches is offset by a smaller sampling area, which is in the range of $100 \sim 400$ μm^2 (Nikooienejad et al., 2020; Marchesi et al., 2021; Sobola et al., 2017), not to mention the need of considerable expertise to obtain high quality images (Kim et al., 2022). As a different approach, deep learning augmentation has been applied as a post-data processing tool to achieve higher resolution from a lower resolution topography, achieving up to ten-fold decrease in sampling time (Kim et al., 2022; Liu et al., 2018). While demonstrating eminent improvement in throughput, such approach still relies on the actual operation of an AFM, bearing an absolute upper limit for throughput improvement. While coupling OM and AFM as a unified system attempted to surpass this intrinsic limitation by initially characterizing regions of interest with OM followed by subsequent operation of AFM, fast scrutiny of only selective regions is made possible. In addition, AFM is inoperable in extreme thermal environments, an additional limitation that has been less visited compared to improving measurement throughput.

Herein, we propose a deep learning-based method that predicts nanoscale topography from conventional OM images, improving throughput (equivalent to OM frame rate, up to 200 FPS) over large probing area ($\sim$1 mm^2) and allowing operation in extreme thermal environments. A multi-domain (OM and AFM) autoencoder is trained to perform three different tasks: 1. predicting nanoscale topography from OM images, and competence on this task allows further analysis of topography. Namely, 2. quantitatively clustering and classifying OM images based on their morphologies, and 3. simulating transformation of a dynamic topography. We demonstrate the feasibility of our proposed methodology using Germanium-on-Nothing (GON) structures that form nanoscale sub-surface cavities upon high-temperature (890°C) vacuum annealing which are utilized in various applications: solar cells (Park et al., 2019), sensors (Kim et al., 2016; Kim et al., 2022), and more (Wong et al., 2019). During annealing, GON structures undergo transformation in surface topography and sub-surface cavities' morphology. For the purpose of topography prediction, the model is trained to learn the OM-AFM correlations of disparate GON surfaces at multiple annealing times, achieving prediction accuracy of about 15% with maximum lateral pixel resolution of 145 nm with the aid of a transfer learning scheme. Based on such quantified relationship, GON structure's OM images are classified based on their morphologies. Lastly, we also realize an empirical simulation of GON's surface transformation over annealing. Unlike many works that utilize deep learning to improve resolution and quantitatively analyze images of a single domain: OM (Wang et al., 2019), AFM (Kim et al., 2022), SEM (de Haan et al., 2019), helium ion microscopes (HIM) (Mill et al., 2021), our work is the first application that quantifies an OM-AFM correlation, therefore predicting nanoscale topographies from OM images.
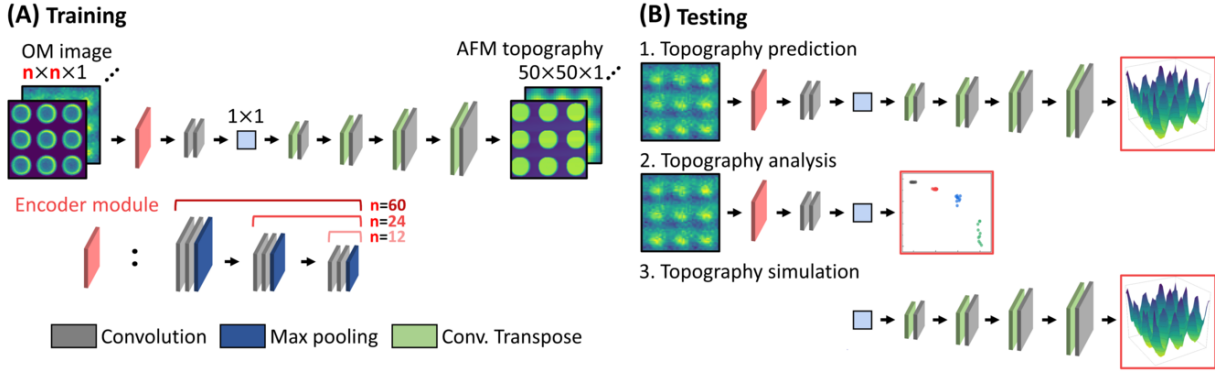
Figure 1: (A) Illustration of the multi-domain autoencoder's architecture, comprised of fully convolutional encoding and decoding to output AFM topography from OM image. (B) Schematic of architectures used for testing. Both encoder and decoder are used for AFM topography prediction, encoder for topography's morphology analysis, and decoder for simulation of topography's morphological transformation.

Methodology

Multi-domain autoencoder

Figure 1 shows the fully-convolutional, multi-domain (OM and AFM) autoencoder model used for this work, which is comprised of the encoder and decoder architectures. The encoder first compressses the OM image into a single latent variable. Then, the latent variable is reconstructed to an AFM topograpghy by the decoder network. OM images with different resolutions and image sizes were used as input data, requiring a tailored encoder for each size. For this reason, the encoder is comprised of a modifiable encoder module and two subsequent convolutional layers. For a 60 by 60 OM image with 100 nm pixel resolution, the encoder module is comprised of 3 convolution blocks, where 2 convolutional layers and a single max pooling layer form a single block of convolution. Encoder modules for OM images of lower resolution are comprised of less convolution blocks, with the lowest resolution of 6 by 6 encoded only using the last two convolutional layers. The decoder network is comprised of three convolution transpose blocks, where each block constitutes one convolutional layer and one convolutional transpose layer. The same decoder architecture is used for all input OM image resolutions since the size of output AFM topography remains identical. All convolutional layers are zero-padded to output an array with the same size as that of the input. To limit the degree of freedom of the autoencoder model, maximum of 10 channels are used for all convolutional layers. The maximum number of channels was chosen in consideration of the complexity of this work's data.

After training, the autoencoder network is capable of three different tasks: 1. AFM topography prediction, 2. quantitative morphology analysis based on OM images, and 3. simulation of dynamic topography transformation. Each task utilizes different parts of the autoencoder architecture. For AFM topography prediction, the entire autoencoder architecture is used to encode the test OM image and reconstruct the AFM topography. For morphology analysis, only the encoder is used to encode the OM images into a single latent variable to cluster and classify them. This variable represents the specimen image's morphological pattern, namely GON's topography for this work. Its magnitude indicates the degree of GON's annealing, enabling monitoring of GON's annealing fabrication process solely with the latent variable from OM images. Lastly, only the trained decoder is used for simulation of AFM topography transformation where the decoder reconstructs AFM topographies from an arbitrary value of latent variable. For our work, the surface transformation of a GON structure upon thermal fabrication is simulated.
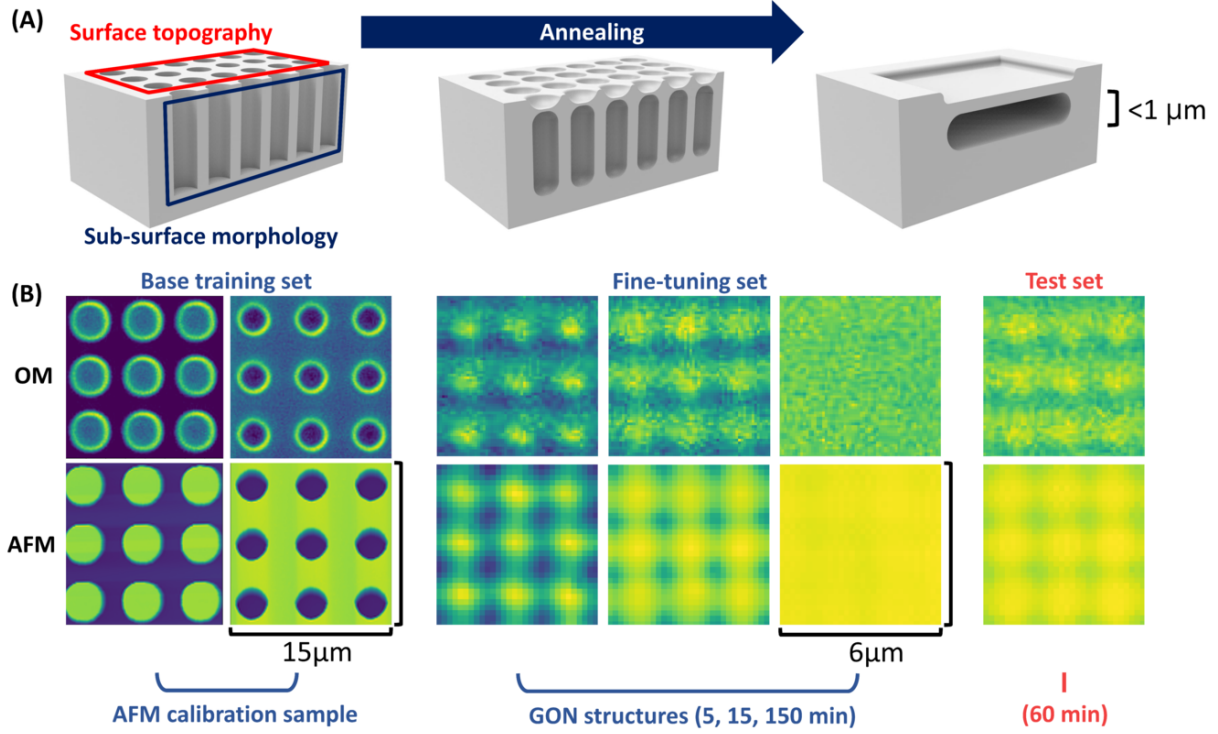
3

Figure 2: (A) Illustrated transformation of surface and sub-surface topography of GON during annealing. (B) Pairs of OM and AFM surface images for base training and fine-tuning. Base training set is acquired from calibration samples, and fine-tuning set from GON structures.

GON fabrication and dataset construction

Figure 2 shows the transformation of surface and sub-surface morphologies of GON structures during the annealing process. In figure 2A, as the initial DRIE pattern is annealed, the surface encloses and a sub-surface single cavity forms due to the surface diffusion at high temperature. For this work, hole patterns were fabricated on the surface of a prime-grade Czochralski (100) germanium (Ge) wafer. The diameter and spacing between hole patterns in figure 2A are 1.2 µm and 0.8 µm, respectively. The Ge hole patterns were annealed in a high vacuum furnace (2×10^{-6} Torr) at 890°C to accelerate the surface diffusion (Jeong et al., 2022). Then, four structures were sampled after different annealing durations: 5, 15, 60, and 150 minutes. Surface image and topography of each structure were respectively acquired using OM and AFM, as shown in the correlations in figure 2B. The hole patterns of annealed GON surface topography were sampled with a 3 by 3 lattice window to incorporate both sufficient robustness against over-fitting and maximal performance of the autoencoder neural network. OM images were acquired with resolution between 6 by 6 and 60 by 60 for a 3 by 3 lattice surface DRIE hole pattern. AFM topographies were acquired with resolution of 30 by 30. While a higher resolution AFM topography is available for acquisition, 30 by 30 has sufficed for proving the feasibility of this work's concept. The latter sub-section further discusses the relationship between OM and AFM resolutions.

4

Transfer learning

Transfer learning, a concept widely used for machine learning where knowledge of a given domain is transferred to a different domain for deeper understanding, has proved its competence for computer vision tasks (Shin et al., 2016). Instead of training a model from scratch for a new domain, fine-tuning a trained model on a similar task has shown superior performance in terms of final accuracy and training efficiency. To utilize such benefit, AFM calibration samples' OM-AFM correlations are also acquired and used for base training, followed by fine-tuning on GON samples. MikroMasch's TGXYZ calibration grating of different step heights (20, 100, and 500 nm) are used for base training, each comprised of four different patterns: two circular and two linear, each pair with inverted heights. Since GON structures manifest a repeating pattern of circular hills, only the circular patterns of calibration samples are initially trained, using two pairs for each step height. After base training, GON structure correlations of different annealing times are used for fine-tuning with a lower learning rate of 0.00075, a 25% lower learning rate compared to 0.001 for base training. For GON structures, 10 OM/AFM pairs were each sampled from 4 GON structures, each with different annealing time: 5, 15, 60, and 150 minutes. Entire 40 pairs of all annealing times are used when constructing a full database, and 30 pairs from three annealing times when constructing a partial database to predict the correlation of the unused annealing time pairs. The remaining 10 pairs are equally split into validation and test set. Both base training and fine-tuning phases train the same autoencoder architecture (figure 1A and 1B) with an Adam optimizer with mean squared error as the loss function. For partial database construction, base training is conducted for 150 epochs with training loss minimization, and fine-tuning for 450 epochs with validation loss minimization. Training loss takes topographies of training set as labels, and validation loss with those of validation set. Training loss is minimized for both base training and fine-tuning when constructing a full database. When training without a transfer learning scheme, the randomly initialized autoencoder model is directly trained on the GON structures dataset with a learning rate of 0.001 for 450 epochs.
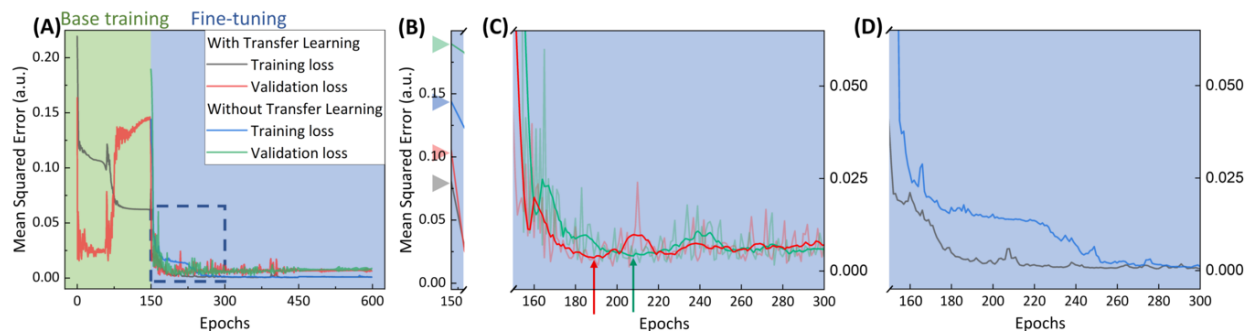

Figure 3: Analysis of transfer learning performance with 60 min. GON structure as test sample. (A) Full comparison of training loss and validation loss for both with and without transfer learning. (B) Comparison of initial training losses and validation losses with and without transfer learning (fine-tuning) (C) Comparison of validation losses, with bold denoting the smoothed validation loss. With transfer learning, validation loss reaches a lower minimum about 20 epochs faster. (D) Comparison of training losses. Training loss converges to its minimum 2.5× faster with a transfer learning scheme.

Figure 3 shows one of the best training results of the autoencoder model when constructing a partial database, with 3 annealing time structures (5, 15, and 150 min.) as training set and the remaining 60 min. structure as validation and test set. In figure 3A, base training is highlighted in green, and subsequent fine-tuning in blue. For both regions, loss represents the mean squared pixel-wise error between the predicted and label topographies. Therefore, an accurate topography prediction incorporates a lower loss value. The model is glorot-initialized where neural net weights are randomly sampled from a normal distribution of

5

the number of input and output units (Glorot & Bengio, 2010). During base training in the green region, validation loss initially decreases as the model learns the circular morphology of AFM calibration samples, which is considerably similar to the GON structures' morphology. However, after about 50 epochs, the model is over-fit to the planar circles of AFM calibration samples and validation error increases due to the discrepancy between the planar calibration samples and non-planar GON structures.

After base training, the same model is fine-tuned on GON structures. Figures 3B to 3D compare the performance of subsequent fine-tuning with and without a transfer learning scheme. The two main advantages of using transfer learning are faster convergence to minimum loss and lower minimum loss value (Pan & Yang, 2009), which is manifested from the training of this work. Figure 3B compares the initial training and validation losses when starting the training on GON structures. Without transfer learning, the initial training (blue line) and validation losses (green) are each nearly double the value of training (grey) and validation loss (red) with transfer learning. Due to such lower magnitude of initial losses, the autoencoder model trained with a transfer learning scheme reaches the minimum validation loss about 20 epochs earlier compared to without transfer learning, as shown in figure 3C. Even with a 25% smaller learning rate during fine-tuning compared to training without transfer learning, the latter requires 50% more epochs until reaching its minimum validation loss. Such an efficient training scheme with transfer learning is even more clearly exhibited when comparing the training losses, shown in figure 3D. Convergence to minimum training loss requires about 2.5 times more epochs without transfer learning. Note that the validation error without transfer learning has reached its minimum about 50 epochs earlier compared to training error, which could be attributed to the discrepancy between training and test data. Excessive training results in over-fitting, where the model is over-optimized to the trend of the training data. This trend is not necessarily congruous to that of test data, therefore yielding a higher validation error with further training beyond the optimal point as the model excessively learns the unique trend of training data. In terms of validation loss magnitude, validation loss without transfer learning converges to a 30.9% higher error value compared to that with transfer learning. The higher validation error resulted in 1.6% higher topography prediction error compared to 7.9% with transfer learning. Although the improvement could not be considered majorly compelling, the latter section explains the improvement in training process stability which proves the significance of transfer learning scheme. Further mentioned training iterations in the discussion section utilize the transfer learning scheme.
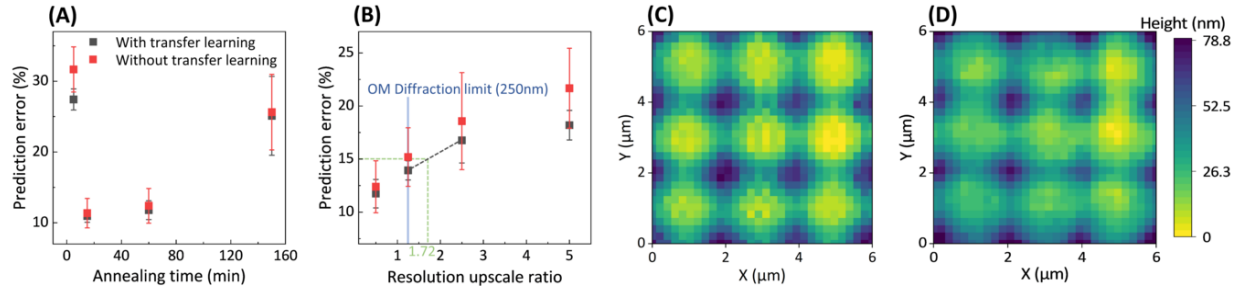

Figure 4: (A) AFM topography prediction error for each annealing time structures. The error bars denote standard deviation. (B) Prediction error for different degrees of resolution upscale from OM to AFM. Upscale degree larger than 1 denotes improvement in resolution. AFM's resolution is fixed as 30 by 30 for all iterations. (C-D) Comparison of 60 min. annealed GON structure's true surface (C) and predicted (D) AFM topography. The illustrated prediction bears 7.9% average error compared to its true surface.

Results & discussion

Topography prediction and its parametric study

One primary parameter that influence the model's topography prediction performance is the test data selection from the dataset of 4 discrete annealing periods. Each annealing time sample is iterated as test data and its topography is predicted based on three other annealing time structures' OM-AFM correlation. The prediction performance varied depending on which annealing time data is selected as test data. **Figure 4** A shows the percentage errors of each annealing time as test data, with and without transfer learning. Each square data point shown in figure 4A and 4B represents the average percentage error of 5 separately trained predictions, with the error bar denoting their standard deviation. Prediction error is calculated by dividing the pixel-wise difference between predicted height and training data height by maximum of the two, followed by averaging of all pixels' errors. As shown in figure 4A, the prediction errors are considerably lower for 15 and 60 min. annealed structures as test data compared to those with 5 and 150 minutes'. For 15 and 60 min. structures, the AFM topography is predicted based on inference of the non-linear trend from both directions in terms of annealing time gradient. For 5 and 150 min. structures as test data, however, the AFM topographies are predicted based on the trend only from a single direction. Specifically, 5 min. structures are predicted based on the decreasing trend of 15, 60, and 150 min. structures, and 150 min. structures based on an increasing trend of 5, 15, and 60 min. structures. Such discrepancy resulted in a drastic difference in prediction performance between 15, 60 min. and 5, 150 min. samples. The remaining analysis of this section is conducted with 60 min. topography as test data.

The quality of topography prediction also depends on the input OM image resolution. The predicted AFM topography's resolution is consistent as 30 by 30 with 200 nm pixel resolution, while OM image resolution varied from 60 by 60 to 6 by 6 which corresponds to resolution upscale from 0.5 to 5. Figure 4B shows the prediction error for different resolution upscale from OM to AFM, which increases as the upscale magnitude increases. With insufficient surface topography information from OM images at lower magnifications, the prediction error soars up to nearly 20% with transfer learning, quantifying the intuitive need for a high-resolution input data. Since reconstructed topograhpies have higher resolution than the abbe diffraction limit of 250 nm, our deep learning-based methodology has shown feasibility in sub-wavelength imaging using a conventional optical microscope. Under a closer scrutiny, an interesting characteristic is observed near the OM diffraction limit of 250 nm, namely at 1.25 resolution upscale ratio in figure 4B. Due to the intrinsic optical diffraction limitation, improvement in OM resolution over 250 nm only picks up mere details of the diffraction pattern and does not provide new useful information. These details could be considered as unique diffractive patterns for each annealing time of training data, in other words noise prone to overfitting. Indeed, improvement in prediction accuracy along with increase in OM resolution conspicuously decreases over 250 nm compared to the trend below 250 nm. When interpolating the prediction error trend between resolution upscale ratio of 1.2 and 2.5, 15% error corresponds to upscale ratio of 1.72. In other words, the maximal achievable resolution upscale within error magnitude of 15% is 1.72. Considering the OM diffraction limit of 250 nm, 1.72 upscale ratio corresponds to 145 nm lateral pixel resolution for the predicted topography, which is the ultimate resolution limit of this work's method.

Figure 4D shows the most accurate prediction among 5 iterations, bearing 7.9% average prediction error. 6.23 nm is acquired as the height error when applying the 7.9% error to the maximum height of predicted AFM topography shown in figure 4D. Considering the few nanometer-range uncertainty of AFM for structures with height less than 250 nm (Klapetek et al., 2011), 6.23 nm prediction error is comparable to AFM's intrinsic measurement error and thus could be considered accurate. The previously mentioned 15% prediction error is about 3-4 times greater than this intrinsic measurement error, which is an acceptable degree considering the significant improvement in throughput (equivalent to OM acquisition rate) and measurement area ($\sim$1 mm^2). These advantages could be merited by any other application as long as sufficient training data is provided for the specific domain, in which this work has dealt with topographies of GON sturctures. Supplementary figure S1 shows topography predictions for all annealing times.

Figure 4A and 4B also quantitatively compares the performances of models trained with and without the transfer learning scheme. For all data points of both figures, the model trained with transfer learning scheme shows slightly lower average prediction error along with an evidently lower standard deviation. Despite the evident trend of lower error with transfer learning, especially for comparisons between different OM magnifications shown in figure 4B, prediction errors for all data points with and without transfer learning intersect to a large degree when considering standard deviation. Therefore, it would be ambitious to claim an evident superiority in AFM topography prediction performance with transfer learning. Nevertheless, the value of transfer learning is manifested with lower standard deviation. Since the autoencoder model with transfer learning first learns the correlations of AFM calibration samples, the model is already established a firm understanding of the OM and AFM correlation for circular patterned structures before encountering GON samples. Subsequent fine-tuning with the presence of such knowledge reduces the randomness of the model's training, therefore significantly lowering the standard deviation in addition to a slightly reduced average error. Given that such merits of transfer learning have been acquired with a dataset of less than 50 samples, faster training and stable performance of the transfer learned model would be even more valued when training on a more complex domain that requires a more rigorous training procedure.
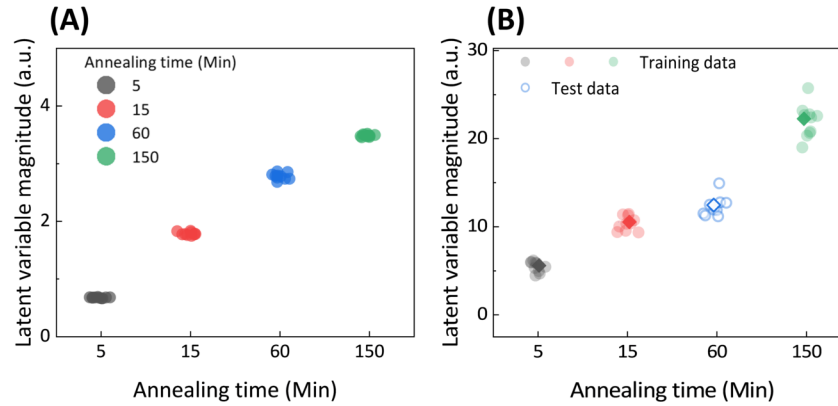

Figure 5: (A) Classification database with all 4 annealing time structures used as training set. The single latent variable is used as the sole parameter for classifying OM images, and therefore its physical implication is analogous to annealing time. All structures are discretely classified based on their annealing time. (B) Classification databases with 60 min. annealed structure as test data. Squares denote the averaged latent variable for each annealing time cluster, and hollow symbols are used for test data. The averaged latent variable of test data (hollow square) is accurately projected between 15 and 150 min. clusters.

Topography classification

In addition to AFM topography prediction, encoder of the same autoencoder model is used for classifying specimen's topography based on its morphology. In doing so, the trained encoder is used to express OM image into a single latent variable based on its topographic morphology, which are then analyzed via clustering and classification. For this work, latent variable infers the degree of annealing for GON fabrication process since the difference in OM images' morphology is due to annealing. Indeed, annealing time and latent variable magnitude show evident correlation as shown in **figure 5** A. Longer annealing duration flattens out the GON structure's surface as shown in figure 2, a trend of morphological transformation precisely represented by the magnitude of latent variable. Such classification is utilized for fabrication monitoring of GON structures. Upon annealing, morphologies of GON's sub-surface cavity and surface topography transform with correlation. Therefore, to manufacture a GON with specific cavity shape in need, the topography is monitored to determine the degree of annealing. Then, the sub-surface structure is predicted based on an empirical database of surface-to-sub-surface morphology correlation over annealing (Jeong et

al., 2022). For the purpose of fabrication monitoring, all 4 annealing time samples are used to construct the full database, where all latent variables have been accurately clustered according to their annealing times (figure 5A). With the constructed database as reference, new GON structures are determined their degree of saturation based on the encoded latent variable from OM image. Utilizing OM's small form factor and versatility in thermodynamic and vacuum conditions (Hayashi et al., 2015), our proposed method is especially useful under extreme thermal conditions where an *in-situ* inspection is unavailable, a common situation for semiconductor manufacturing. Our real-time monitoring scheme could be applied to all fabrication and manufacturing processes that require a quantitative analysis of surface pattern and its transformation.

Figure 5B demonstrates an application of the proposed fabrication monitoring methodology, using the same trained model used to predict the topography shown in figure 4D. While the database of figure 5A is constructed using structures of all annealing times, only three annealing times (5, 15, and 150 min.) are used as training data for constructing the database shown in figure 5B. The remaining 60 min. annealing time structure is used as test data to predict its degree of saturation compared to the three clusters of training data. Here, unlike the full database with discrete clusters for all annealing times, the 60 min. test structure is not discretely clustered, with its range's half overlapping with that of the 15 min. cluster. To resolve such error, each cluster's 10 data points are averaged as denoted by the bold diamonds in figure 5B, resulting in comprehensive representations. In doing so, the 60 min. annealed structure's averaged latent variable is accurately projected between the 15 and 150 min. annealed structures' latent variables. Note that the Y -scale range varies between those of full database (figure 5A) and partial database (figure 5B). Such discrepancy is not only due to the difference in training data but also the randomness incorporated in deep learning model's initialization. Even under identical training conditions and data, the absolute range of latent variables varies by each iteration while the overall distributions of latent variables remain analogous.
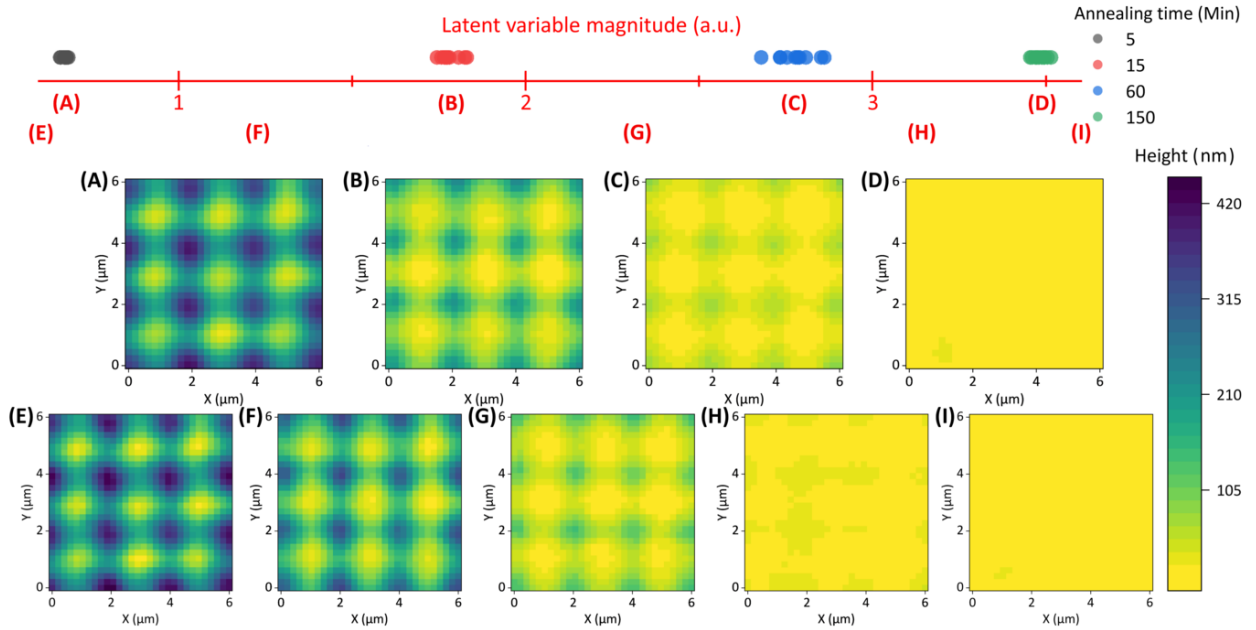

Figure 6: GON structures' AFM topography simulation using the trained decoder. The top 1-dimensional graph expresses the magnitude of the encoded latent variable. Encoded distributions of training data are plotted above the 1-dimensional axis with colored circles and arbitrary latent variable magnitudes for topography reconstruction are plotted below the graph as alphabets. Topographies of figure 6A-D are generated from the region where training data are distributed, and 6E-I are generated in-between the distributions of training data.

Simulation of topography transformation

Figure 5A's full database accurately classifies the OM images using a single latent variable encoded by the encoder. Compared to the encoder, the roll of a decoder is to reconstruct AFM topographies from latent variables. If a latent variable is arbitrarily sampled from the range of encoded training data, the decoder will reconstruct a topography based on the correlations of training data. In that sense, if the training data incorporates a temporal transformation of a physical phenomenon, as in this work, the decoder is capable of simulating the overall process. **Figure 6** precisely shows such simulated results of the temporal transformation of GON structure's topography. The x -axis at the top denotes latent variable magnitude. Colored circles above the x -axis represent the encoded latent variables of the training OM images as shown in figure 5A, and the red letters below the axis represent the latent variable magnitudes that are sampled for topography reconstruction. Among the red letters, the upper row (a) through (d) are sampled from the range of training data's latent variable distributions for each annealing time. Therefore, these generated topographies are nearly identical to their respective training data topographies, with an average percentage error of 3.27%. The lower row, (e) through (i), are sampled from region where training data are not present. (f) through (h) are sampled at the middle between training data clusters, while (e) and (i) are sampled before 5 min. and after 150 min. clusters, respectively. Supplementary video (**Figure 7**) dynamically elaborates such simulation with 1 minute time steps.

Figure 7: Simulation of GON structures' topography transformation with sampling interval of 1 minute

Since figure 6E to 6I are sampled from where training data are not available, these are reconstructed based on the topography transformation trend of training data, namely figure 2B. Overall, as figure 6 illustrates, the planarizing process of GON structures' topography is competently characterized. Specifically, the decoder reconstructed topographies not only within the training data's latent variable range but also beyond each endpoint. Although figure 6E and I are reconstructed beyond each former and latter endpoints, both reasonably depict the surface morphologies based on the transformative trend between 5 and 150 min. surfaces. In reality, however, despite their intuitive and reasonable topographic simulations in these extreme latent regions, physically fabricated structures could result in a different topography at these temporal regions. While the continuously flattening trend after 150 minutes is accurately accounted for, simulated trend before 5 minutes is different to its actual topographic morphologies. Initial phases of annealing transformation start from DRIE cylindrical holes shown in figure 2A. These holes are initially etched at the yellow regions of figure 6A, and therefore the heights at these regions would need to approach to a flat zero plane instead of an hill as annealing time decreases down to zero. Had OM-AFM correlations of the initial pattern been also added as training data, the decoder will be able to characterize the topographic transformation between initial pattern and 5 minutes of annealing. Therefore, despite the qualitatively reasonable reconstruction of topographies outside the training data's domain, only the reconstructions within the latent variable range of training data would be considered physically accurate. An autoencoder-based simulation or reconstruction of any other domain would also entail such limitation, emphasizing the importance of acquiring more training data to extend the applicable scope and improve its validity.
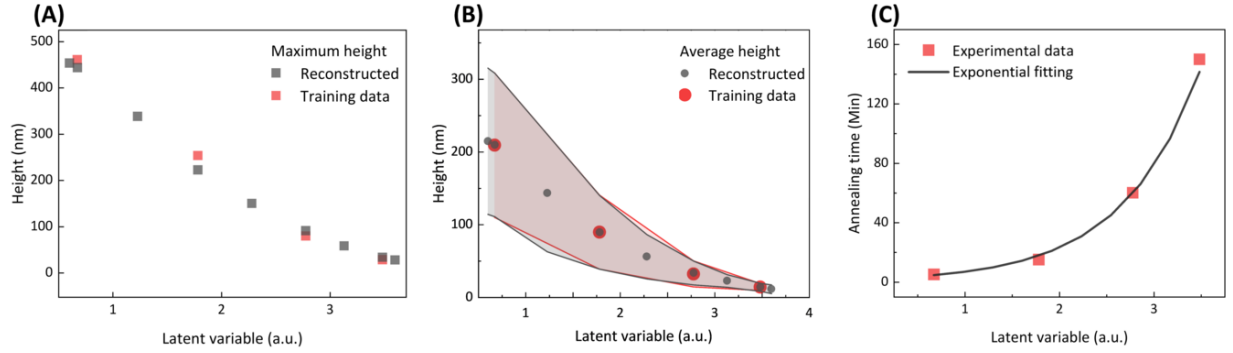
Figure 8: Quantitative analysis of trained decoder's AFM topography reconstruction. (A) Comparison of maximum height of each training data and reconstructed topography. (B) Comparison of average height of each training data and reconstructed topography. Error bands denote the standard deviation of each topography's height. (C) Scattered and fitted correlation between latent variable magnitude and annealing time. An exponential relationship between the two is shown, fitted by $y=1.994e^{1.223x}$.

While figure 6 demonstrates the decoder's reasonable reconstruction of surface topography transformation, a quantitative analysis would further validate the simulation accuracy. In that sense, **figure 8** A and 8B compares the topography's maximum height, average height, and standard deviation between training data and reconstructed topographies, which are three key features that characterize topographies. As latent variable increases, its corresponding topography flattens out in a non-linear fashion, characterized by the decreasing maximum and average height in figure 8A and 8B. In addition, as represented by the error bar of figure 8B, standard deviation also decreases over increasing latent variable magnitude due to flattening of surface upon annealing. The reconstructed topographies (grey) accurately follow all three non-linear trends established by training data (red), quantitatively proving the competence of trained decoder's simulation.

Under a close scrutiny, error between maximum height of training data and reconstructed topographies of figure 8A is found to be greater than that of average height of figure 8B. Indeed, unlike figure 8A, the red datapoints of training data in figure 8B completely overlay the corresponding grey datapoints at same latent variable magnitudes. Such discrepancy is due to the nature of how our autoencoder model is trained. The autoencoder model has been trained to minimize the mean squared height error between the training data and reconstructed topographies. Therefore, the model has trained to be competent in reconstructing topographies with overall resemblance, which results in small difference between training data and reconstructed topographies' average heights. In comparison, the maximum height of a topography is determined only by height of a single pixel and therefore is prone to a relatively larger error. Nevertheless, the fact that the reconstructed topographies' maximum heights follow the training data's non-linearly decreasing trend proves the competency of this work's proposed decoder-based simulation methodology. Compared to conventional simulation methodologies such as finite element method or phase-field modeling (Zhang et al., 2017), the beauty of our proposed methodology is its simplicity. The autoencoder model does not require any modeling of governing physical rules since the neural net autonomously learns the pattern incorporated in the training data. Therefore, our method allows for an efficient and accurate empirical simulation without any knowledge of specimen material or its physical characteristics.

The trained decoder's simulation shown in figure 6 illustrates the reconstructed topographies from sampled latent variable magnitudes. However, the latent variable itself does not bear a physical meaning, therefore limiting the applicability of the proposed simulation framework. Without any link between a physical property and the simulated data, interpreting and comparing the simulated data with experimental data is inevitably limited to a superficial comparison. For our work dealing with temporally transforming surface

11

topography, the important parameter for the simulated data is time. Therefore, a quantitative link between latent variable magnitude and time is vital for practical use of the proposed simulation results. Such crucial need has been realized by fitting the 2D plot of latent variable and annealing time, as shown in figure 8C. Their relationship is competently represented by an exponential fit with a R-squared value of 0.9933 using the following equation: $y = 1.994e^{1.223x}$. Utilizing such relationship, this work's simulated results are able to be quantitatively interpreted in reference to a physical parameter. For an even improved accuracy, the encoder could be selectively trained to obtain a tailored latent space distribution as explained in the supplementary materials, where the R-squared value is acquired as 0.9999. The proposed simulation method is applicable regardless of physical domain, as long as the relationship between latent variable and a physical parameter is accurately represented as in this work.

Overall, this work's GON structures manifest a considerably simple morphology of flattening circular peaks and therefore a single latent variable-based architecture has sufficed for the autoencoder model. Recently, the same architecture has been proven its competence in dealing with more complex images for traditional computer vision tasks such as scene segmentation (Badrinarayanan et al., 2017) and human face generation (Hou et al., 2017). In that sense, based on this work's proven feasibility of an autoencoder model-based simulation of a physical phenomenon based on images of transforming morphology, the same autoencoder-based simulation would also be applicable to more complex multi-domain phenomena (Makin et al., 2020; Yang et al., 2021), an extensive possibility to be further explored in the future.

Conclusion

This work proposes a multi-domain autoencoder framework that predicts nanoscale topography from OM images. We show the feasibility of this deep learning-based methodology using GON structures, achieving prediction error as low as 7.9% with 1.25× upscale in lateral resolution. Using the same framework, OM images are quantitatively classified based on their morphologies, and the dynamic transformation of topography is empirically simulated. The novel OM-based approach is operable under extreme conditions, along with improvement in topography measurement throughput (up to 200 FPS) and area($\sim$1 mm^2). Such competence has been realized with the use of transfer learning, which is fine-tunable to any specific type of topography. Therefore, we envisage the use of the multi task-oriented autoencoder in various scientific and engineering domains that require high-throughput analysis of topography and its morphology.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

References

Nanorheology of living cells measured by AFM-based force–distance curves. (2020). *Nanoscale*, *12*(16), 9133–9143. https://doi.org/10.1039/c9nr10316c

Nanoscale detection of ionizing radiation damage to DNA by atomic force microscopy. (2008). *Small*, *4*(2), 288–294.

Indications of chemical bond contrast in AFM images of a hydrogen-terminated silicon surface. (2017). *Nat. Commun.*, *8*(1). https://doi.org/10.1038/ncomms14222

Atomic fingerprinting of heteroatoms using noncontact atomic force microscopy. (2021). *Small*, *17*(51), 2102977.

Giant Enhancement of Spin-Orbit Torque Efficiency in Pt/Co Bilayers by Inserting a WSe 2 under Layer. (2021). *Adv. Electron. Mater.*, *8*(1), 2100684. https://doi.org/10.1002/aelm.202100684

High-throughput atomic force microscopes operating in parallel. (2017). *Rev. Sci. Instrum.*, *88*(3), 033703. https://doi.org/10.1063/1.4978285

AFM-Based nanofabrication and quality inspection of three-dimensional nanotemplates for soft lithography. (2021). *J. Manuf. Process.*, *66*, 565–573. https://doi.org/10.1016/j.jmapro.2021.04.051

Video-Rate Non-Raster AFM Imaging With Cycloid Trajectory. (2020). *IEEE Trans. Control Syst. Technol.*, *28*(2), 436–447. https://doi.org/10.1109/tcst.2018.2879939

An ultra-wide scanner for large-area high-speed atomic force microscopy with megapixel resolution. (2021). *Sci. Rep.*, *11*(1). https://doi.org/10.1038/s41598-021-92365-y

Influence of scanning rate on quality of AFM image: Study of surface statistical metrics. (2017). *Microsc. Res. Tech.*, *80*(12), 1328–1336.

Accelerating AFM Characterization via Deep-Learning-Based Image Super-Resolution. (2022). *Small*, *18*(3), 2103779.

General Resolution Enhancement Method in Atomic Force Microscopy Using Deep Learning. (2018). *Adv. Theory Simul.*, *2*(2), 1800137. https://doi.org/10.1002/adts.201800137

Germanium-on-Nothing for Epitaxial Liftoff of GaAs Solar Cells. (2019). *Joule*, *3*(7), 1782–1793. https://doi.org/10.1016/j.joule.2019.05.013

Hollow Microtube Resonators via Silicon Self-Assembly toward Subattogram Mass Sensing Applications. (2016). *Nano Lett.*, *16*(3), 1537–1545. https://doi.org/10.1021/acs.nanolett.5b03703

Self-assembled silicon membrane resonator for high vacuum pressure sensing. (2022). *Vacuum*, 111101. https://doi.org/10.1016/j.vacuum.2022.111101

Extended Design Space of Silicon-on-Nothing MEMS. (2019). *J. Microelectromech. Syst.*, *28*(5), 850–858. https://doi.org/10.1109/jmems.2019.2927466

Deep learning enables cross-modality super-resolution in fluorescence microscopy. (2019). *Nat. Methods*, *16*(1), 103–110.

Resolution enhancement in scanning electron microscopy using deep learning. (2019). *Sci. Rep.*, *9*(1), 1–7.

Synthetic image rendering solves annotation problem in deep learning nanoparticle segmentation. (2021). *Small Methods*, *5*(7), 2100223.

Thick Germanium-on-Nothing Structures by Annealing Microscale Hole Arrays With Straight Sidewall Profiles. (2022). *J. Microelectromech. Syst.*, *31*(2), 183–185. https://doi.org/10.1109/jmems.2021.3139094

Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. (2016). *IEEE Trans. Med. Imaging*, *35*(5), 1285–1298. https://doi.org/10.1109/tmi.2016.2528162

Understanding the difficulty of training deep feedforward neural networks. (2010). *Aistats*, *9*, 249–256. https://proceedings.mlr.press/v9/glorot10a.html

A survey on transfer learning. (2009). *IEEE Trans. Knowl. Data Eng.*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Methods for determining and processing 3D errors and uncertainties for AFM data analysis. (2011). *Meas. Sci. Technol.*, *22*(2), 025501. https://doi.org/10.1088/0957-0233/22/2/025501

PCA-based sub-surface structure and defect analysis for germanium-on-nothing using nanoscale surface topography. (2022). *Sci. Rep.*, *12*(1). https://doi.org/10.1038/s41598-022-11185-w

High Temperature Microscope Observation of Melt Formation at the Interface between Crystalline Olivine and Wüstite. (2015). *ISIJ Int.*, *55*(6), 1210–1212. https://doi.org/10.2355/isijinternational.55.1210

A study of morphological evolution of silicon microstructure based on phase field model. (2017). *Ferroelectrics*, *520*(1), 154–158. https://doi.org/10.1080/00150193.2017.1375319

SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. (2017). *IEEE Trans. Pattern Anal. Mach. Intell.*, *39*(12), 2481–2495. https://doi.org/10.1109/tpami.2016.2644615

Deep Feature Consistent Variational Autoencoder. (2017, March). *2017 IEEE Winter Conf. Appl. Comput. Vis.*. https://doi.org/10.1109/wacv.2017.131

Machine translation of cortical activity to text with an encoder–decoder framework. (2020). *Nat. Neurosci.*, *23*(4), 575–582. https://doi.org/10.1038/s41593-020-0608-8

Multi-domain translation between single-cell imaging and sequencing data using autoencoders. (2021). *Nat. Commun.*, *12*(1). https://doi.org/10.1038/s41467-020-20249-2