Deep learning-based prediction of fabrication-process-induced structural variations in nanophotonic devices

Gostimirovic, Dusan; Xu, Dan-Xia; Liboiron-Ladouceur, Odile; Grinberg, Yuri

Publisher's version / Version de l'éditeur:
https://doi.org/10.1021/acsphotonics.1c01973
ACS Photonics, pp. A-K, 2022-07-20

ACS Photonics

pubs.acs.org/journal/apchd5

Article

Deep Learning-Based Prediction of Fabrication-Process-Induced Structural Variations in Nanophotonic Devices

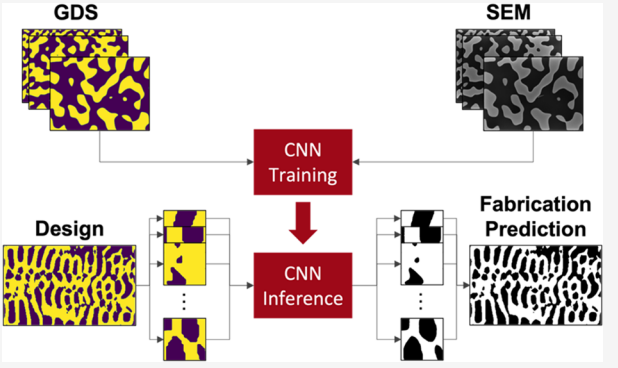
Dusan Gostimirovic,* Dan-Xia Xu, Odile Liboiron-Ladouceur, and Yuri Grinberg

Cite This: https://doi.org/10.1021/acsphotonics.1c01973

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The performance of integrated silicon photonic devices is sensitive to small structural variations that arise from imperfections in the nanofabrication process. This sensitivity is exacerbated for next-generation devices that require fine feature sizes to push the limits of performance. In this work, we present a deep convolutional neural network model to predict fabrication variations in planar silicon photonic devices and verify their manufacturing feasibility prior to prototyping. Our model is trained on a modest set of scanning electron microscope images of structures that experience dimensional inaccuracies stemming from combined contributions from proximity effects in lithography and loading effects in dry etching. Our model quickly and accurately predicts over/under-etching, corner rounding, filling of narrow channels and holes, and washing away of small features in a photonic device. With this, the expected performance of a device can be predicted through an extra simulation and any necessary design corrections can be made prior to fabrication.



KEYWORDS: *silicon photonics, integrated photonics, machine learning, deep convolutional neural networks, fabrication process variations, topological optimization*

INTRODUCTION

Integrated silicon photonic circuits are expected to enable a future of all-optical and optoelectronic computing and telecommunications with low-loss, low-power, and high-bandwidth capabilities—all without significantly changing the existing microelectronics fabrication infrastructure.[1,2] Higher levels of performance are achieved through modern design elements such as subwavelength grating metamaterials[3,4] and inverse-designed, topologically optimized structural patterns[5,6] that push the feature sizes of the nanofabrication technology to its limits. Although these devices show high performance under ideal simulation conditions, they often perform differently in experiment.[7] Highly dispersive devices like vertical grating couplers, wavelength (de)multiplexers, and microresonators can experience significant performance deviation from just a few nanometers of structural variation caused by imperfections in the nanofabrication process.

By restricting the minimum feature size of a device, certain designs can work within a reasonable target performance range. For sensitive, dispersive devices, where a few nanometers of wavelength shift can completely change the intended operation, restricting minimum feature size is not enough to guarantee good performance, and the device will often require tuning circuitry to "fix" the performance post process, adding to complexity and power consumption. To mitigate fabrication variations by current best practices, a designer will manually calibrate a design by creating a range of pre-biases from the nominal to increase the chances of hitting the desired performance target.[8,9] However, this process is inefficient with time and chip space, scales poorly with design complexity, and leaves the designer with little valuable information to carry over to other designs. In addition, there are significant process variation effects that cause nonuniform over- and under-etching depending on the curvature and proximity of features that cannot be accounted for by uniform pre-biasing. For example, a tight convex curve (silicon inside the curve) will experience over-etching, and a tight concave curve (silicon outside the curve) will experience under etching. Simple, uniform pre-biasing can only account for one of these two variations. Furthermore, the amount of over- or under-etching changes throughout the curve, where the apex generally experiences the most variation. Previous works have demonstrated methods that scale the amount of variation according to random spatially varying manufacturing errors to design devices that are more robust to random errors across the chip and between different fabrication runs,[10] but to the best of our knowledge, there have been no demonstrations of a biasing model that scales the amount of variation with respect to the degree of feature curvature.

Received: December 22, 2021

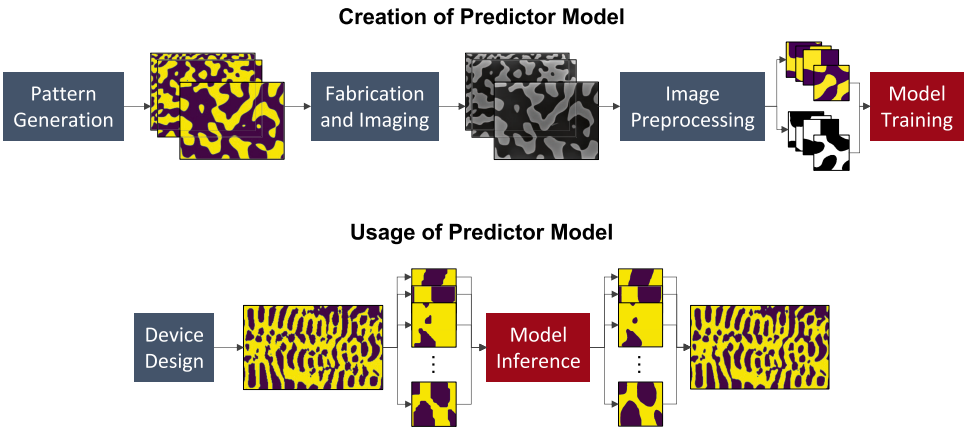


Figure 1. Procedure for the creation of the predictor model is outlined in the top sequence, and the usage of the predictor model is outlined in the bottom sequence.

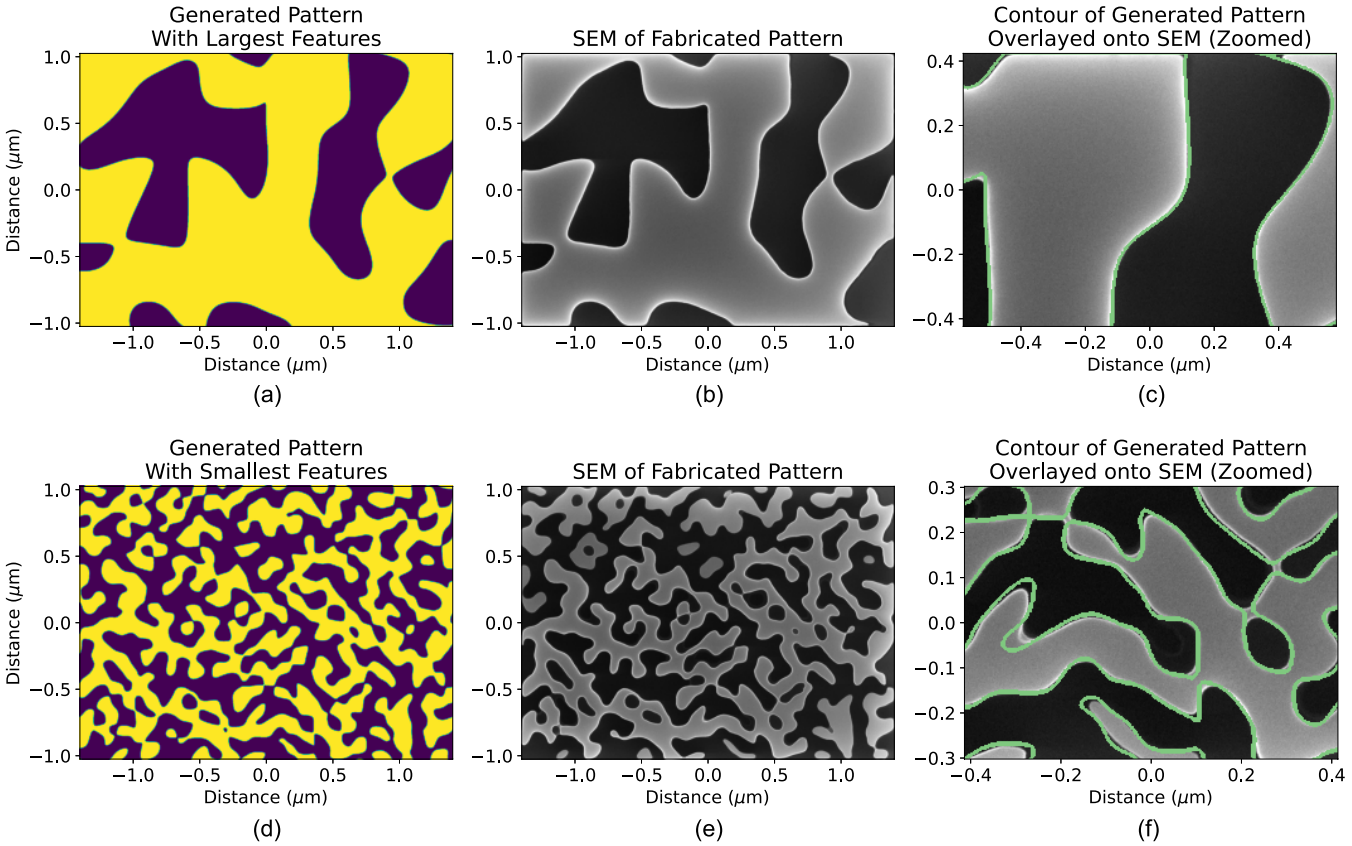


Figure 2. Two examples of the 30 design patterns in the dataset. The top row is of the pattern with the largest average feature size, and the bottom row is of the pattern with the smallest average feature size. (a, d) Generated design patterns (GDS), where yellow regions are silicon and purple regions are silica. (b, e) Corresponding SEM images. (c, f) Zoomed portions of the SEMs with green contours of the GDS overlaid on top to demonstrate fabrication variations.

Lithographic proximity effects cause unintended exposure to nearby areas of a device (commonly seen with the rounding of sharp features),[11] and etch loading effects cause different etch rates for features with differing sizes or differing amounts of surrounding open area.[12] Proprietary tools simulating these effects, which are based on physical models and developed primarily for microelectronics with Manhattan-type geometries, are available to pre-emptively calculate and correct for, e.g., proximity effects;[13−15] however, proper use of these tools requires in-depth knowledge and process-specific calibration that are generally not available to external users (designers). Furthermore, applying the process is not straightforward for nanophotonic devices, which generally contain features of highly varying curvature and pattern density, where even the standard, dose-based proximity effect corrections applied by most foundries cannot perfectly reproduce them. In other words, despite the best effort in state-of-the-art fabrication facilities to improve patterning fidelity, deviations from the nominal design still occur for the finer features, which ultimately degrades device performance. These are serious challenges the silicon photonics community is contending with.[16]

https://doi.org/10.1021/acsphotonics.1c01973
ACS Photonics XXXX, XXX, XXX−XXX

There has been a recent effort to address this problem. For example, purely predictive modeling methodologies[17] based on process-specific physical models that are intended for microelectronics but have only been applied to conventional photonic structures with simple, straight features. In this work, we set out to create a universal fabrication variation predictor, as outlined in Figure 1, that accounts for the complex unwanted effects in the multistep fabrication processes without having to give special consideration to each effect. For this, we propose an ensemble of deep convolutional neural network (CNN) models that are trained on image examples obtained from pairs of graphic design system layouts (GDS) and their corresponding scanning electron microscope (SEM) images. To the best of our knowledge, our work is the first effort in predicting fabrication variations with a purely data-driven approach that is intended for complex, fine-featured (inverse-designed, topologically optimized) photonics devices. We use an integrated photonics foundry using electron-beam lithography, but the methodology can be readily applied to other fabrication technologies such as deep ultraviolet (UV) lithography. The structures from the dataset only take 0.01 mm^2 of chip space and 30 SEM images to fully capture. The images are prepared for training by an automated process and can be readily applied to any nanofabrication process by simply refabricating and reimaging. Although our modeling process is designed for topologically optimized devices, the resulting model also accurately predicts fabrication variations in conventional photonic devices, where the impact is more subtle.

Our goal is 2-fold: first, to use the CNN model as a low-cost replacement for the fabrication and imaging steps in design validation, and second, to identify design features with strong inherent fabrication uncertainty and take measures to minimize their presence. Furthermore, we believe that this work opens the door to future prediction-enhanced optimization algorithms that generate highly robust, high-performance photonics devices with minimal extra computation and fabrication costs. To stimulate further research and discussion on the topic, we have made the code and data readily available in the GitHub repository, PreFab (Prediction of Fabrication).[18]

METHODS

Data Acquisition and Preparation. For machine learning tasks, having data that properly represent the statistical distribution of the ground truth is critical. For this, we use a reputable integrated photonics foundry for nanofabrication and SEM imaging. In addition to needing high-quality data, a large amount of highly varied data is required to train a well-generalized model. Because of the large costs and lengthy timelines associated with nanofabrication, we have designed our training structures in a way that allows us to acquire a large dataset of high-quality training images with minimal chip space and imaging time. Our dataset is populated with 30 randomly generated 2.8 × 2.0 μm^2 patterns (with nanometer resolution) from different Fourier transform-based filters, as shown by the examples in Figure 2 (see the Supporting Information for a detailed description of the pattern generation process). This allows us to generate a large dataset of variable structural features like those of topologically optimized devices without having to design multiple real devices. The pattern size is chosen to fit an SEM image of a desired size and resolution. The generated patterns are not themselves optically useful. The Fourier transform filter size of each pattern determines its average feature size. We use two filter types (low pass and band pass) to increase the variety of features in our dataset, which improves the generalizability of the model. In addition to the main, wavy features, the hard boundaries of the patterns create useful sharp features otherwise not created by the pattern generator. Although our dataset only contains features generated by one method (with slightly varying conditions), it includes many variations of features and feature spacings for the model to generalize well to the types of photonics structures we expect to predict. Should there be a case for predicting specialized devices with different boundary conditions (e.g., devices densely integrated with others; large periodic structures; small, isolated structures), and the current model performs poorly, specialized patterns can be fabricated, imaged, and added to the dataset to improve the capabilities of the model.

The generated patterns were fabricated on a 220 nm silicon-on-insulator (SOI) platform by electron-beam lithography through a silicon photonic multi-project wafer service by Applied Nanotools Inc.[19] As is standard by most foundries, the patterns received a baseline dose-based proximity effect correction to improve pattern fidelity. However, it is evident by Figure 2f that this does not perfectly reproduce the fine features of the original design. After lithography and etching, a 3.0 × 2.25 μm^2 SEM image with a resolution of 1.5 nm/pixel was taken of each pattern. This pattern size strikes a good balance between image quality and the amount of structural data obtained from each image. After fabrication and imaging, the GDS and SEM images are processed to prepare the dataset for training. This process includes binarization of both images, cropping to the edges of the patterns, setting equivalent image resolution, and aligning the images together (see the Supporting Information for a detailed description of the pretraining image processing). The images are then sliced into 128 × 128 $pixel^2$ slices to fill the dataset with a manageable number of variables per training example. The slicing process scans through an image in overlapping steps of 32 pixels. A total of 50,680 slices are obtained. This process of taking data from different image perspectives is known as data augmentation,[20,21] which is a common method of artificially creating more training data. Finally, each GDS slice is matched with its corresponding SEM slice.

Because our dataset is based on a specific set of fabrication processes, the CNN model is specific to this fabrication technology. However, our methodology can be extended to many other scenarios, including other fabrication technologies (e.g., deep UV lithography) or other material platforms (e.g., III−V or different waveguide thicknesses): the generated patterns would simply need to be fabricated and imaged once. Likewise, should the same fabrication process shift over time (i.e., parameters of the tools change through manual adjustments or performance drift), the model can be recalibrated by reimaging and retraining. Given the simplicity and speed of our process, this can be performed regularly to keep the model up to date. The modeling process, in effect, can replace the conventional calibration of process monitoring structures, which only provides information for a simple set of design features.

Training the Predictor Model. In this work, the CNN predictor model is trained to learn the relationship between the GDS input and the corresponding SEM outcome. A CNN works similarly to conventional, fully connected multilayer perceptron neural networks but with additional convolutional

https://doi.org/10.1021/acsphotonics.1c01973
ACS Photonics XXXX, XXX, XXX−XXX

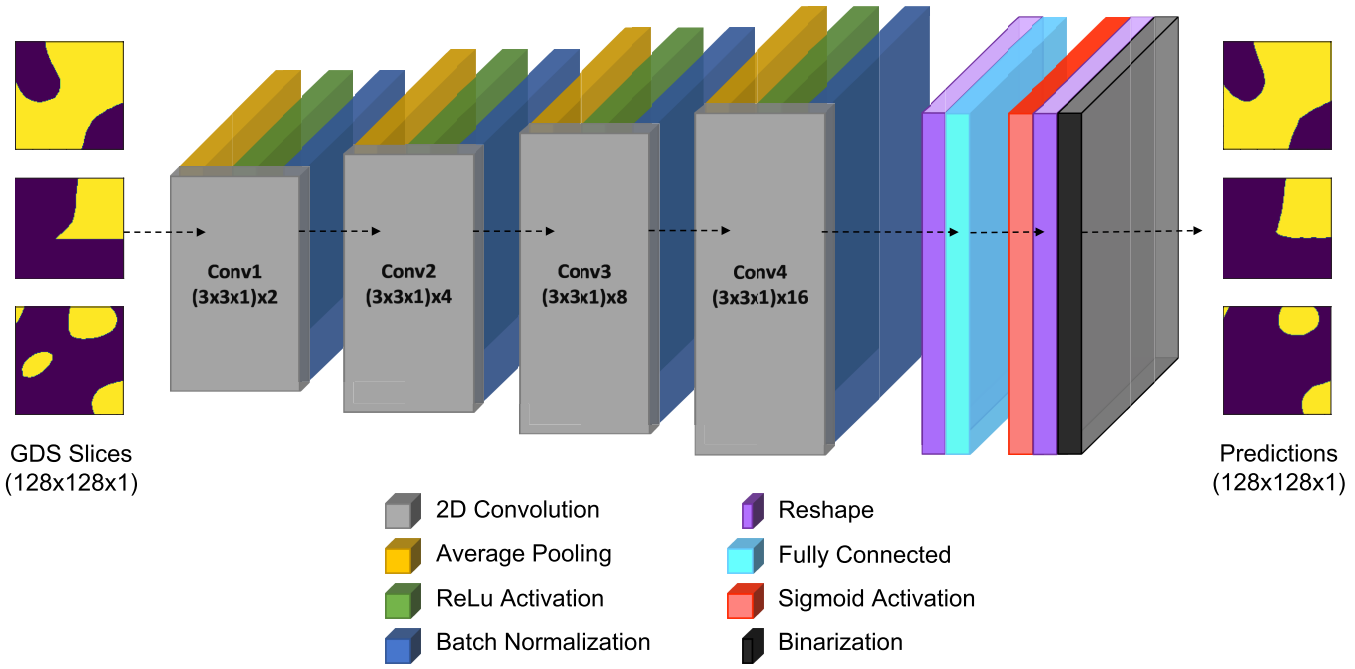


Figure 3. Network structure of a CNN fabrication variation predictor model. Four convolutional layers are followed by a reshape layer that flattens the 2D output of Conv4 to a one-dimensional (1D) array compatible with the fully connected layer. The second reshape layer converts the output back to a 2D image. Note that the binarization layer is only included in inference (not training).

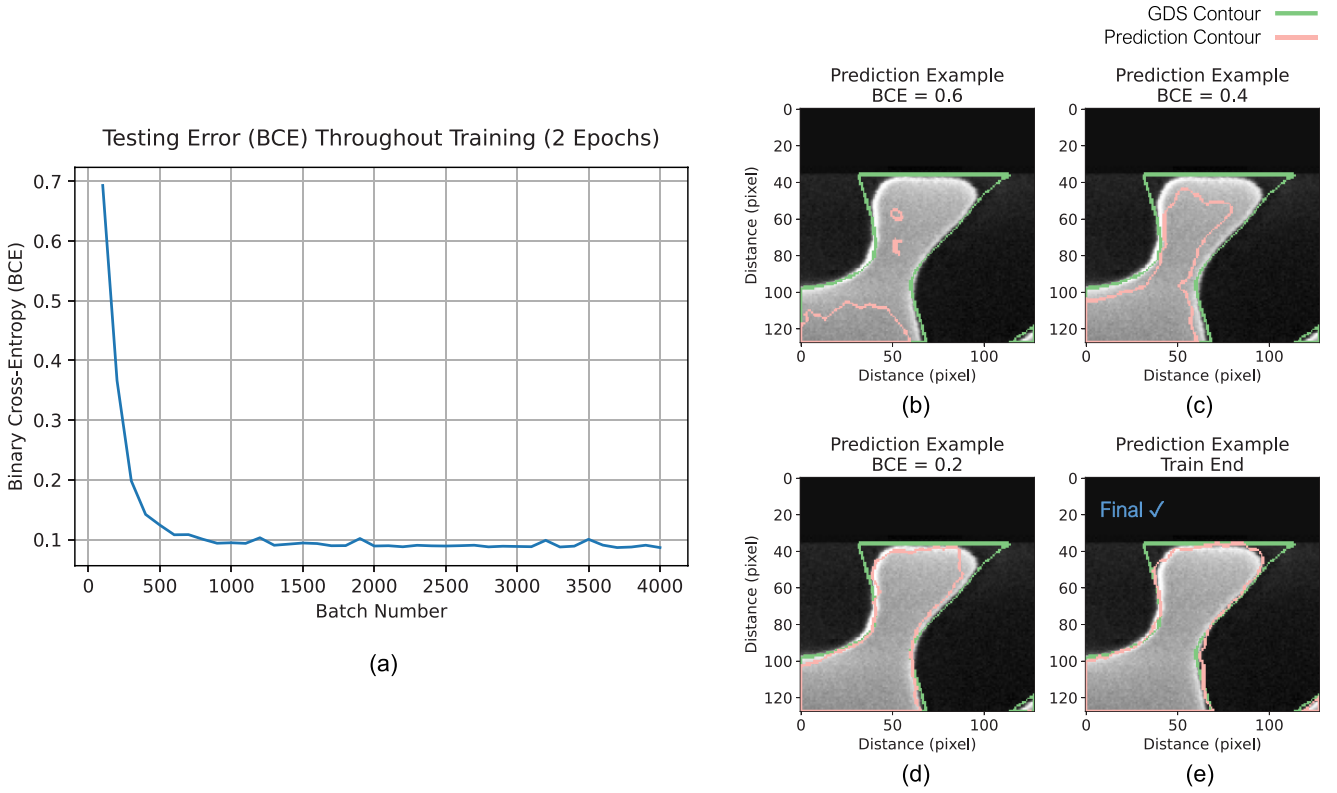


Figure 4. (a) Testing BCE (error) of the CNN predictor model over two epochs. An example SEM slice (from the test dataset) with overlaid GDS and prediction contours at (b) BCE = 0.6, (c) BCE = 0.4, (d) BCE = 0.2, and (e) BCE = 0.085 (final iteration).

layers at the front of the network. The convolutional layers make the CNN more suitable for identifying and classifying complex features in images. These networks conventionally take a full image as an input and classify its contents. Our variation of the CNN still takes an image as the input, but the output is now a matrix of silicon-or-silica (core-or-cladding) classifications based on the learned effects of the nanofabrication process.

Our model is constructed and trained using the open-source machine learning library, TensorFlow,[22] and the detailed setup

https://doi.org/10.1021/acsphotonics.1c01973
ACS Photonics XXXX, XXX, XXX−XXX

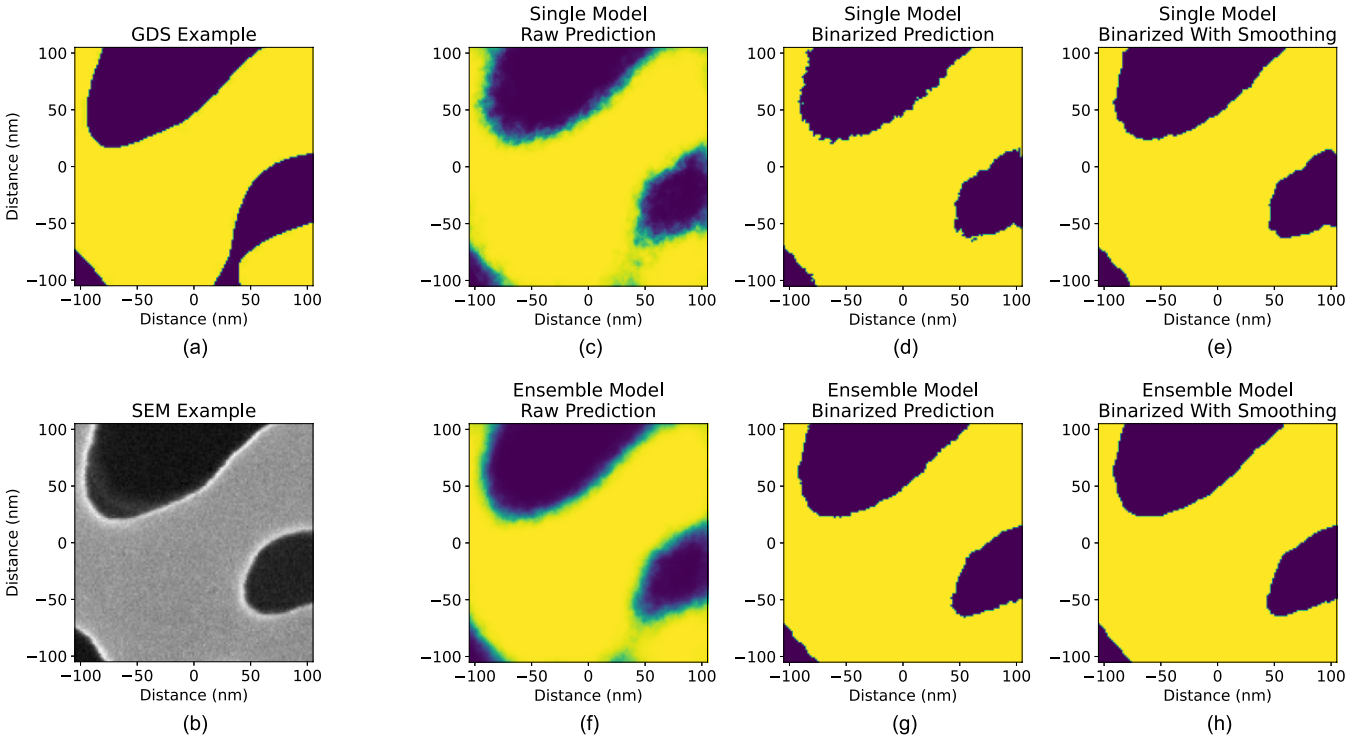


Figure 5. Prediction steps of (a) a single GDS example slice (from the training dataset). (b) Corresponding SEM slice for reference. The example slice is fed into the CNN model to produce (c) the raw prediction. Green pixels indicate uncertainties, where the prediction is between 0 and 1. By (d) directly binarizing the prediction, the edges of the features become rough, so we use (e) a Gaussian blur to smoothen the edges before binarization. (f–h) Corresponding prediction steps for an ensemble model, which averages multiple predictions from multiple models together.

is shown in Figure 3. At the front of the model are four two-dimensional (2D) convolutional layers—each with average pooling to downscale the image, ReLu activation to add nonlinearity, and batch normalization to accelerate learning and improve generalizability. This is a common network arrangement for image classification tasks,[23,24] and it proves to work equally well for our case. The convolutional layers work to detect features (e.g., edges, corners, circles) in our input images (GDS) and relate them to the transformed output images (SEM). By combining multiple convolutional layers in a "deep" arrangement, the network can detect and classify more complex hierarchical patterns related to the specific dataset. At the end of the model is a single fully connected layer with a sigmoid activation to map the convolutions back into a 128 × 128 pixel^2 prediction.

The network weights are trained with the adaptive moment estimation method (Adam) and the binary cross-entropy (BCE) loss function. Adam is a popular, computationally efficient optimizer for image classification tasks,[25,26] and it proves to perform well for our case as well. BCE is useful for binary classifier tasks such as ours. For each pixel of the GDS image, the model classifies the probability of it being silicon: 1 being 100% silicon, 0 being 100% silica, and anything in-between being an uncertainty. For a perfectly trained model, the in-between values are minimized and BCE = 0. While in practice achieving BCE = 0 is not realistic, achieving low errors is possible and often sufficient for the network to perform well.

Prior to training, we split the dataset so that 50% of the data is used for training and 50% is used for testing. With this distribution, each partition receives one full pattern randomization of each filter type/size. The training dataset is randomly shuffled, a batch size of 16 is used, and the training runs for two epochs (where every image in the training set has been fed through the model twice). The BCE progression over training time is shown in Figure 4, where the testing error is shown to be minimized at the end. Figure 4 also shows the prediction of an example in the test dataset at different stages of training to further illustrate the error/accuracy progression. Our network arrangement and combination of parameters lead to a low error of BCE ≈ 0.08; however, better performance may be achieved with further network parameter refinements. Should even higher accuracy be desired, we believe that improving the quality of the data—through higher-resolution imaging and more careful alignment and binarization in the preprocessing stage—will result in the biggest improvement. For this model, adding more of the same kind of data will not make a significant improvement; in fact, we find that we can reduce the size of our training dataset to 10% of the overall dataset size before seeing a significant drop in accuracy.

Making Single Predictions. Using the process in Figure 3, an example is shown in Figure 5 on how a single prediction (of an example in the testing dataset) is made by inputting a 128 × 128 pixel^2 slice of the design and running a forward pass (inference) through the CNN predictor model. One prediction takes approximately 50 ms on a low-power 16-core GPU. A raw prediction, as shown in Figure 5c, is made at the final, fully connected output layer of the CNN. At the raw output, each predicted pixel may be silicon, silica, or somewhere in-between based on the certainty of the model. The in-between (uncertain) values are a result of imperfections in the training setup and random variations in the nanofabrication process. The training imperfections come from suboptimal network parameters and imperfect data. These can both be improved by further refining the network structure and hyperparameters and

https://doi.org/10.1021/acsphotonics.1c01973
ACS Photonics XXXX, XXX, XXX−XXX

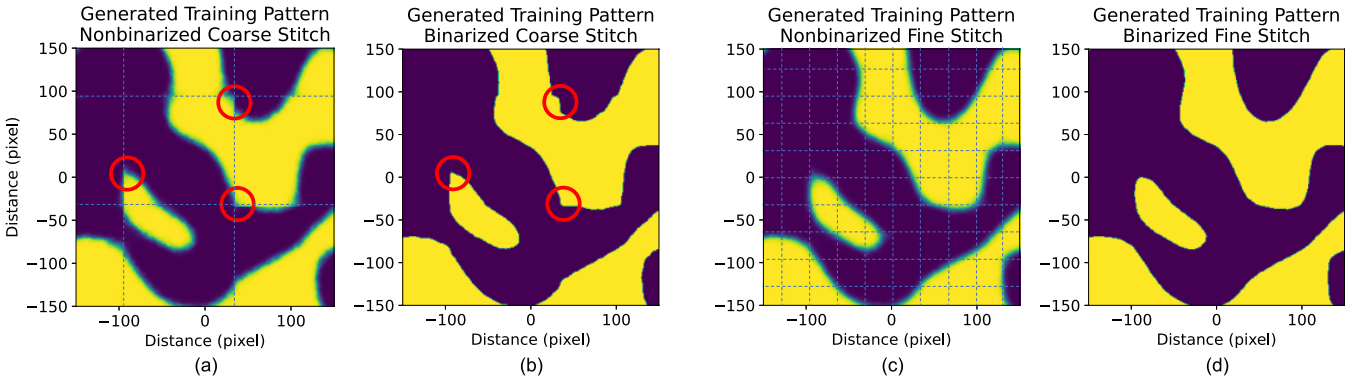


Figure 6. Prediction (zoomed in with overlayed stitch grids for demonstration) of a generated pattern with (a) no binarization and a coarse stitching step size (128 pixel), (b) binarization and a coarse stitching step size, (c) no binarization and a fine stitching step size (32 pixel) with overlap averaging, and (d) binarization and a fine stitching step size with overlap averaging.

increasing the resolution of the images. Random variations in the fabrication process are more difficult to mitigate, though our model represents them well using uncertainty values. These process variations can arise from changes across the surface of the chip (e.g., variation of plasma density in etching, wafer bowing) and small variations in the fabrication equipment over time (e.g., e-beam drift). The resulting random structural variations are more significant for structures near and past the minimum feature size limits. For the example in Figure 5c, there is a narrow channel near the bottom (at $x \approx 50$ nm, $y \approx -100$ nm) that has more uncertain pixels (neither fully yellow nor purple) in the raw prediction. Channels like this occur throughout the dataset, and they can get "bridged" after fabrication, like in this example, but not always. Inconsistencies in the outcome of seemingly similar channels are caused by complex physical factors inherent to the fabrication processes that this network structure cannot easily learn and will therefore result in uncertain predictions. Similarly, small silicon islands may fall over and be washed away in some examples but remain in others—resulting in further difficulties for the model to understand. Although the prediction accuracy tends to be lower for designs with many small features, the prediction uncertainty carries additional value as a rough metric of manufacturing feasibility/reliability and can be a tool for the designer in finding which features to modify. Regardless, a decision must be made for the final prediction and therefore any pixel above or equal to 0.5 will be silicon. To minimize roughness on the uncertain edge areas, a Gaussian blur is applied before binarization (i.e., the Gaussian blur eliminates the high-frequency components created by forced binarization). The size of the Gaussian blur (5×5 pixel^2 kernel) is set to remove the rough pixels without modifying the main features of the structure.

To minimize training uncertainties that show up in the raw prediction due to the network model and to further increase prediction accuracy, an ensemble of models is used to make a final prediction. Ensemble learning is a common approach used to improve the robustness of machine learning models.[26,27] We trained 10 identical models on the same dataset but with different randomized weight initializations and shuffling of training data. Because the optimization of these deep neural networks is highly nonconvex, each instance of the model will likely end up in a different local minimum and will perform slightly differently. By averaging the predictions of all models together, we remove outlying mispredictions of individual pixels and produce a final prediction with fewer uncertain pixels. This is evident by the smoother raw prediction in Figure 5f and the subsequent binarization of it. The average prediction error of the examples in the testing dataset (calculated as mean-squared error) for the ensemble model is 0.5% lower than the best-performing individual model and 6% lower than the worst. Although the improvement is not large on average, the ensemble model achieves consistently low relative error for each prediction example.

Making a Complete Prediction. Since the training data consists of small, 128×128 pixel^2 (211×211 nm^2) slices, predicting the fabrication variation of a full device design requires multiple predictions to be made and stitched together, as shown by the (zoomed) example in Figure 6. Fabrication variations are highly dependent on the physical size of the features, so the device image to be predicted must first be scaled to the resolution of the training images (1.5 nm/pixel). Because the individual image slices often contain partial features at the boundaries (i.e., there is some missing structural context), the accuracy there will often suffer. As a result, the stitched device prediction, termed coarse stitching, can have misalignments and bumps at the seams. We use an improved stitching process, termed fine stitching, with finer "scanning" steps of 32 pixels, that averages overlapping offset predictions, ensuring that each feature can be predicted away from the slicing boundaries to create a smoother and more accurate final prediction. The stitching step size can be further reduced, but the improvements are marginal. The example in Figure 6 is from one of the generated patterns (zoomed in for demonstration), where 1,400 and 21,200 predictions (each prediction being 128×128 pixel^2) were made and stitched together for the coarse and fine stitching methods, respectively. Recall that the prediction of each slice is already made up of predictions from 10 different models. For the coarse stitch in Figure 6a,b, multiple stitching errors are observed (marked by red circles). These stitching errors do not appear for the finer stitch in Figure 6c,d. Using a larger slice size in the CNN training process would reduce stitching issues, but the increased scale would be more computationally expensive to train and may lead to worsened individual prediction accuracy. Despite the low accuracy near the boundaries of each predicted slice, this fine overlapping stitch method produces a smooth prediction for any full structure of any size.

It should be noted that CNNs can in principle be scaled appropriately to predict full devices without the need for stitching. However, as discussed above, there are trade-offs in building a well-trained CNN model. Working with larger

https://doi.org/10.1021/acsphotonics.1c01973
ACS Photonics XXXX, XXX, XXX−XXX

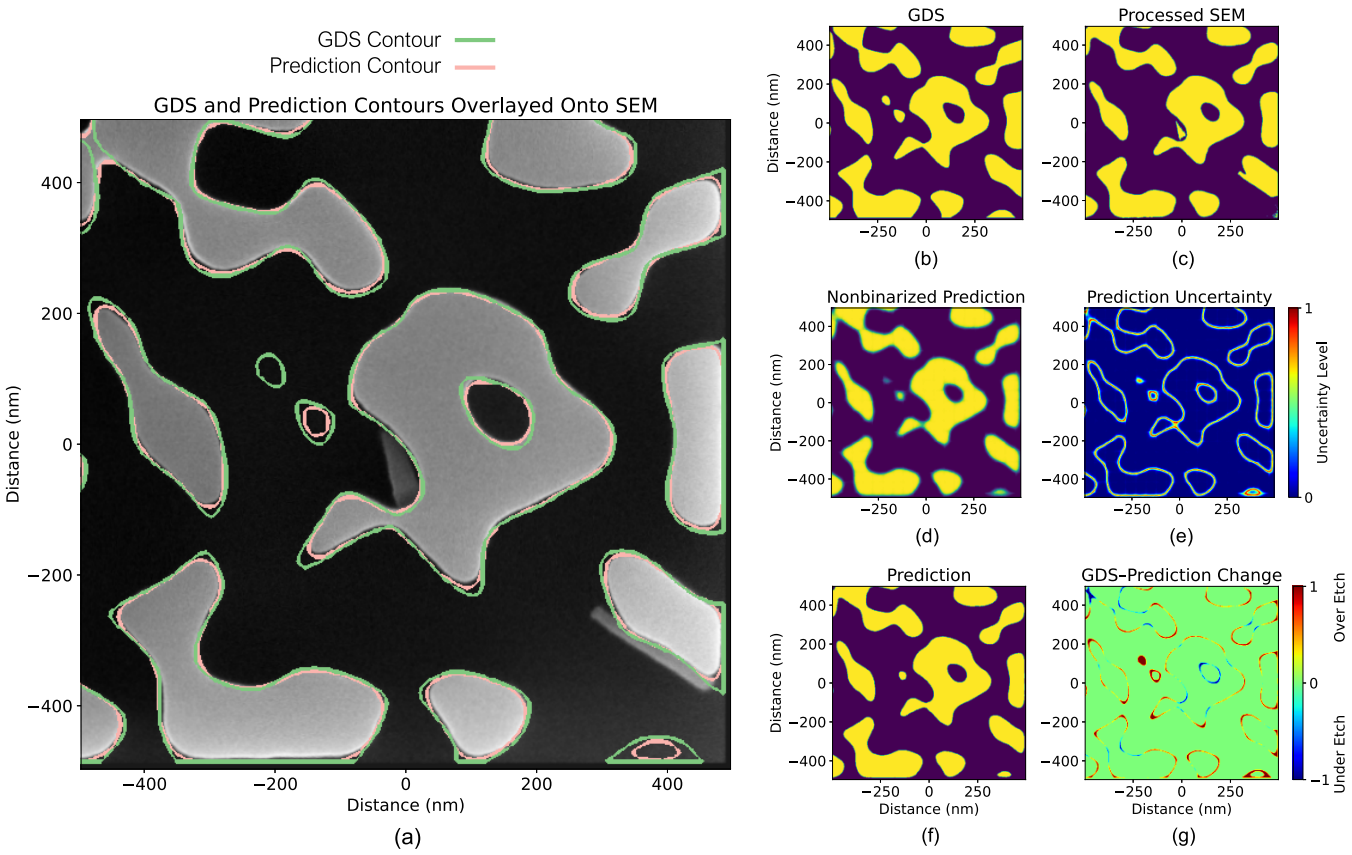


Figure 7. Visual analysis of a generated test pattern (zoomed in for demonstration), which includes: (a) SEM with overlayed GDS and prediction contours, (b) GDS, (c) processed SEM, (d) nonbinarized prediction, (e) uncertainty of the prediction, (f) binarized prediction, and (g) difference between the GDS and the prediction, which shows the distribution of over-etch and under-etch for different feature characteristics.

images/slices (with more pixels/variables) leads to larger networks and higher computation cost.[28] For a given set of SEM images, using larger slices also translates to a dataset with fewer examples. Additionally, to reach close to nanometer resolution for SEM images with a limited frame size (2048 × 1536 pixel^2 in our case), a typical photonic device would require several full SEM images to completely cover, thereby necessitating stitching regardless. The slice−predict−stitch process is also more flexible when predicting devices of different shapes and sizes, as we can build up a full prediction with small, "unit cell" pieces rather than matching the model to one specific device. We believe the stitching strategy we developed addresses these limitations and does not add significant complexity and computation cost (given the millisecond-scale prediction time for each slice).

RESULTS AND DISCUSSION

Figure 7 presents a full prediction example (zoomed in for demonstration) that showcases the capabilities of the CNN predictor model. The example is taken from a generated pattern in the testing dataset and therefore has not been seen by the model in training. Figure 7a presents a zoomed portion of the SEM of the example, with overlaying GDS design and prediction contours for comparison. The longer, straighter edges in the example do not experience much fabrication variation, other than a slight over-etch. The design and SEM differ more greatly where the design has tight bends and corners. At these points, proximity effects cause unequal exposure and rounding, which the CNN model predicts. From Figure 7g, it is further evident that the model does more than uniformly shift the silicon boundaries (like that of a conventional biasing method).[8] For long, straight sections, the over-etching is uniform but increases with higher degrees of convex bending (silicon inside the curve). For concave bends (silicon outside the curve), the silicon experiences under-etching. This is verified by the close overlap between the SEM and the prediction contour for bends and corners in Figure 7a. The certainty of the degree of over/under-etching is limited by the resolution of the SEM images (1.5 nm/pixel), which is evident by a thin uncertainty line remaining in Figure 7e (defined as $U = 1 - 2|0.5 - y|$, where y is the prediction image/matrix). This uncertainty line increases in magnitude for features smaller than the process-specified feature size limit and drops to zero in larger areas of only silicon or silica.

The model sometimes makes mispredictions if small islands remain standing, like those at ($x \approx -200$ nm, $y \approx 100$ nm) and ($x \approx 400$ nm, $y \approx -500$ nm). These small features are affected more by proximity effects and experience additional over-etching due to having no surrounding silicon to protect them. The model can predict the high degree of over-etching for these islands; however, being isolated also means these islands have less structural support and tend to get washed away. Islands near the process-specified minimum feature size may or may not get washed away in the resist removal stage, which the model cannot accurately predict. This is evident by the high degree of uncertainty for the small islands in Figure 7e. If the pixel value is slightly higher than 0.5, the model will keep the island through binarization, but the likelihood of it washing

https://doi.org/10.1021/acsphotonics.1c01973
ACS Photonics XXXX, XXX, XXX−XXX

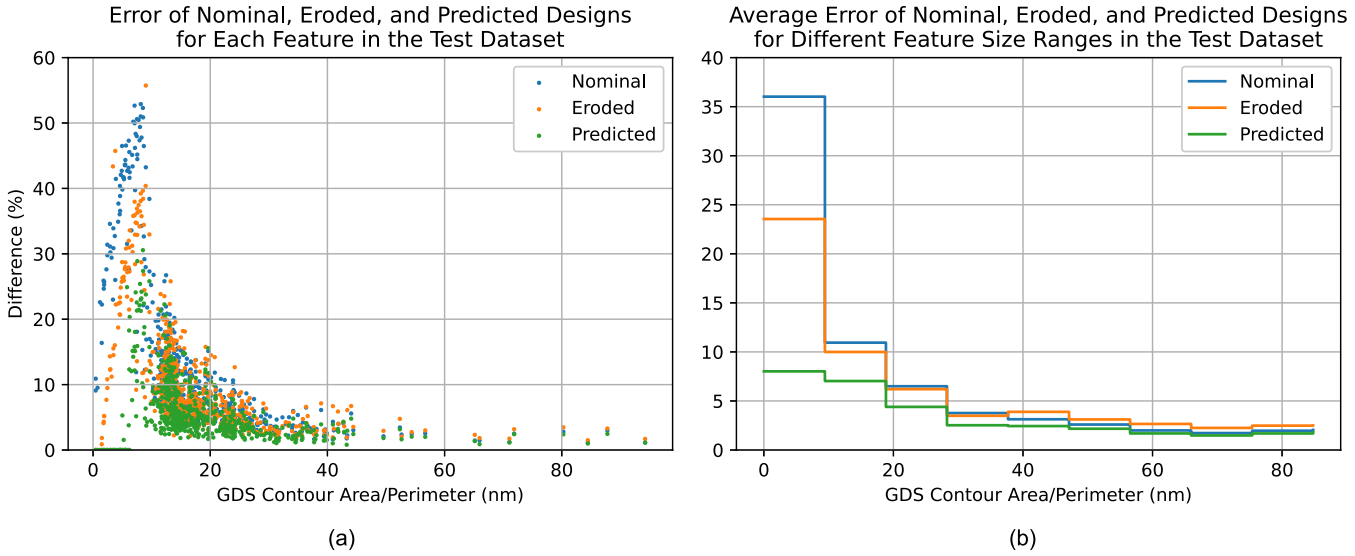


Figure 8. (a) Prediction−SEM, eroded-GDS−SEM, and GDS−SEM differences for each contour in the test dataset of images as a function of contour size (measured as contour area divided by contour length). (b) Averaged prediction−SEM, eroded-GDS−SEM, and GDS−SEM differences as a function of feature size.

away is still relatively high. Islands much smaller than the minimum feature size are easier for the model to predict, as they will more likely be washed away. For any feature with high uncertainty, it is advised that the designer take measures against it (e.g., designing for larger, simpler features with less prediction uncertainty).

The example in Figure 7 also illustrates more specialized capabilities of the predictor model over conventional methods, with, e.g., the filling of a narrow channel at ($x \approx -500$ nm, $y \approx 500$ nm). With a uniform bias, these gaps would be widened, but in fabrication, they get filled due to proximity effects and the difficulty to fully etch through narrow resist openings. This also occurs for small holes in the dataset (not shown in the Figure 7 example). The fallen features at ($x \approx 400$ nm, $y \approx -300$ nm) and ($x \approx 0$ nm, $y \approx 0$ nm) provide further insight into the capabilities of the model. These fallen features sometimes get picked up by the SEM processing/binarization step, as shown in Figure 7c. These fallen features not appearing in the final prediction demonstrates the high generalization of the trained model, as it is learning the physical process effects rather than outputting directly what it has seen.

One potential limitation of our approach is that the relatively small slicing window does not capture much of the surrounding structure and therefore cannot take longer-range proximity effects directly into account. Given that these effects can be corrected for by the baseline, dose-based proximity effect correction process, we believe they are small in comparison to the residual short-range variations (e.g., corner rounding, washing away of small features) we see and that our model predicts. However, a future work that captures more of the surrounding structure in the images may help to further reduce the small uncertainties our model shows in this analysis.

Figure 8 presents the differences in prediction−SEM and GDS−SEM for each structural feature in the training dataset (15 patterns). For each generated pattern, a full fine-stitched prediction is made. After prediction, the contour of each structural feature is extracted, and the percentage of unequal pixels between prediction and SEM, and GDS and SEM, are calculated. For an additional comparison, we add eroded versions of the GDS contours to represent that of a commonly used uniform over-etch bias. The degree of erosion, approximately 3 nm, is chosen to give the best fit to the test dataset (lowest average error). The differences are plotted in Figure 8a as a function of contour area divided by contour perimeter, which increases as features get larger and less complex. For 93% of the features, the prediction is closer to the SEM than the eroded design is. This is expected, as the uniform bias does not capture the varying degrees of over/under-etching (demonstrated by our model in Figure 7g), bridging of narrow channels, and the washing away of small islands. As the design features become smaller and more complex, a larger discrepancy between SEM and GDS is observed, as expected. In these cases, the prediction still aligns rather well with the SEM contours, which indicates its greater benefit for complex, fine-featured, topologically optimized devices. This is demonstrated in Figure 8b, which averages differences for 10 equally spaced contour size ranges. Note that the slight leveling in prediction−SEM difference for the smallest contour size range is because the predictor sees more small islands that get washed away: in these cases, the prediction and SEM are both empty, which leads to many simple, perfect predictions. For complex, fine-featured devices, we believe the predictor model can serve as a more valuable guide than adhering to an absolute feature size limit, as some features larger than the limit may vary significantly and some features below the limit may still be feasible to keep.

To demonstrate how well the model generalizes to patterns outside of the testing dataset, Figure 9 presents the prediction of two different grating couplers containing subwavelength grating (SWG) structures[29] and a topologically optimized wavelength demultiplexer (DEMUX).[30] These devices were fabricated using the same process but on separate runs. The two gratings contain many sharp corners that get rounded in prediction and fabrication. The various sharp features at the boundaries of the patterns in our dataset allow the model to accurately predict more conventional, Manhattan-like structures like these. The topologically optimized DEMUX contains features more like those of our training and testing datasets. In this example, we see the filling of a hole at ($x \approx 0$ nm, $y \approx -250$ nm), the other major variation in this example being the

https://doi.org/10.1021/acsphotonics.1c01973
ACS Photonics XXXX, XXX, XXX−XXX

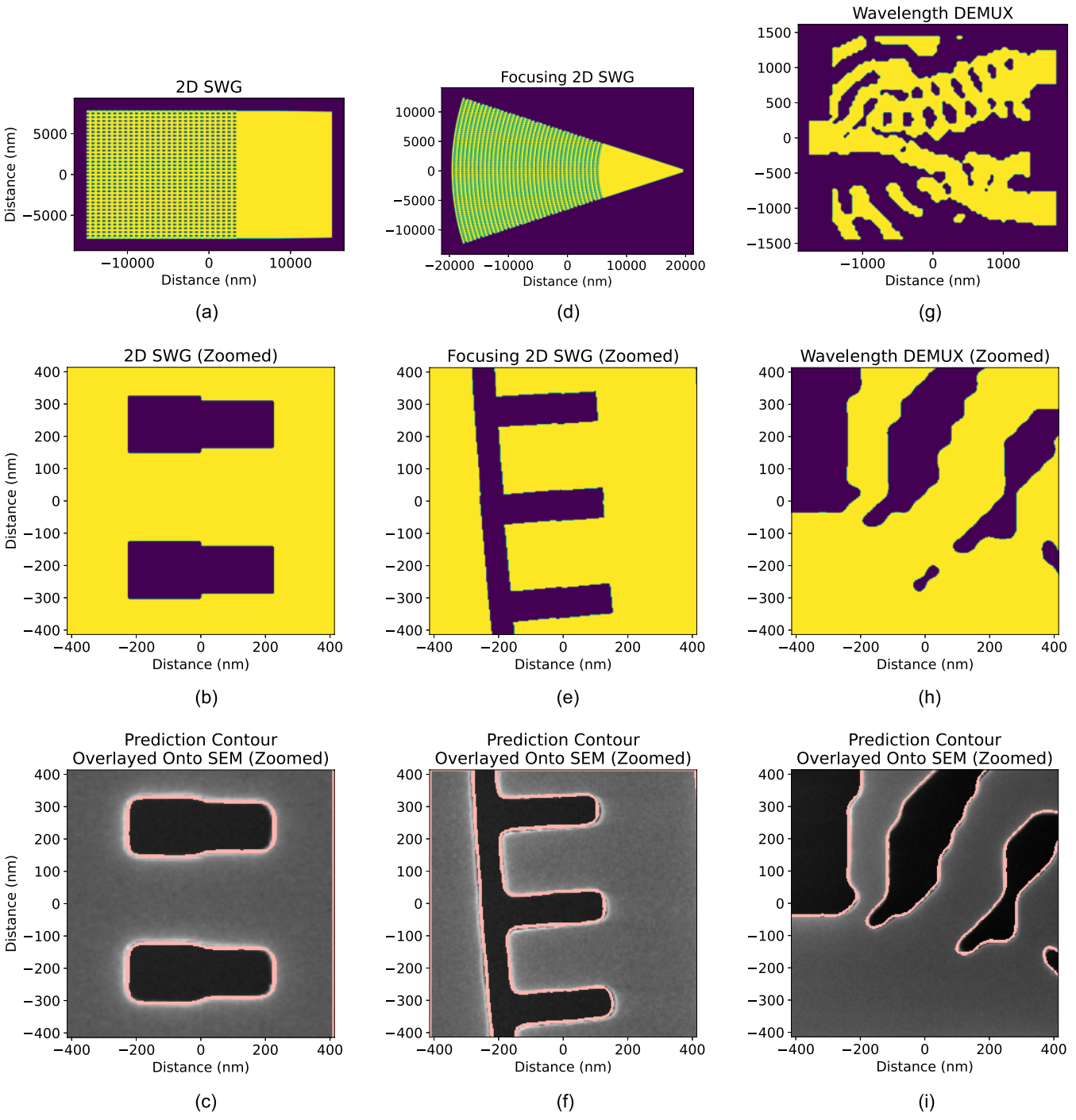


Figure 9. (a) Grating coupler with subwavelength structures (connecting waveguide excluded), (b) zoomed portion of it, and (c) zoomed SEM image of it with an overlayed prediction contour. Corresponding images for (d–f) focusing grating coupler and (g–i) topologically optimized wavelength DEMUX.

smoothing of the pixelated edges of the device. Despite these types of edges not being directly included in the training dataset, the model is able to interpret them and make an accurate prediction. In comparison to the uniform erosion method, our model achieves 40, 37, and 32% reductions in error (calculated as mean-squared error) for the three devices in Figure 9, respectively.

CONCLUSIONS

Inherent physical limitations such as proximity effects, as well as process variations, such as etch-rate dependence on pattern density in nanophotonic device fabrication, can cause severe performance degradation and delayed prototyping. Despite best efforts in state-of-the-art fabrication facilities, such variations are still inevitable. In this work, we present a deep convolutional neural network model for the prediction of planar fabrication variations in silicon photonic devices and propose it as a tool to validate the feasibility of a design prior to fabrication and further offer the possibility of pre-lithography correction. These capabilities reduce the need to perform multiple correctional fabrication runs and accelerate the prototyping of nanophotonic devices and circuits,

https://doi.org/10.1021/acsphotonics.1c01973
ACS Photonics XXXX, XXX, XXX−XXX

representing significant savings in cost and time. In other words, the CNN model can serve as a surrogate for design validation for a particular fabrication technology. The CNN training data is constructed from a modest set of SEM images. We demonstrate how our model quickly and accurately predicts the fabrication variations in a wide distribution of structural features. As features become smaller and more complex, as is typical for highly sensitive topologically optimized devices, fabrication variations become more significant and vary markedly with feature curvature and density. The relative benefit of our predictor model is increased for these types of variations, as simple, linear biasing fails to represent them. Compared to current process simulation tools, including proximity effect calculation and correction, our approach is entirely data-driven; knowledge of the processing specifics and material parameters, which is typically not available to photonics designers, is not required here. The current model is based on an electron-beam lithography process, but the same methodology can be directly applied to other fabrication technologies (e.g., deep UV lithography). For the next-generation photonics devices, this tool allows the designer to know a priori how far to push the size and complexity of the features. We envision future design methods to integrate fabrication variation predictor models into the optimization algorithm to automate the creation of truly robust and high-performing photonic devices.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsphotonics.1c01973.

Fourier transform-based random pattern generator; image preprocessing for CNN training (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Dusan Gostimirovic − *Department of Electrical and Computer Engineering, McGill University, Montreal H3A 0G4, Canada;* orcid.org/0000-0001-9323-4452; Email: dusan.gostimirovic@mcgill.ca

Authors

Dan-Xia Xu − *Advanced Electronics and Photonics Research Centre, National Research Council Canada, Ottawa K1A 0R6, Canada*

Odile Liboiron-Ladouceur − *Department of Electrical and Computer Engineering, McGill University, Montreal H3A 0G4, Canada*

Yuri Grinberg − *Digital Technologies Research Centre, National Research Council Canada, Ottawa K1A 0R6, Canada*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsphotonics.1c01973

Funding

The authors received funding from the National Research Council Canada Grants AI4D-101-1 and HSTN-219.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Siew, S. Y.; Li, B.; Gao, F.; Zheng, H. Y.; Zhang, W.; Guo, P.; Xie, S. W.; Song, A.; Dong, B.; Luo, L. W.; Li, C.; Luo, X.; Lo, G.-Q. Review of Silicon Photonics Technology and Platform Development. *J. Lightwave Technol.* **2021**, *39*, 4374−4389.

(2) Thomson, D.; Zilkie, A.; Bowers, J. E.; Komljenovic, T.; Reed, G. T.; Vivien, L.; Marris-Morini, D.; Cassan, E.; Virot, L.; Fédéli, J.-M.; Hartmann, J.-M.; Schmid, J. H.; Xu, D.-X.; Boeuf, F.; O'Brien, P.; Mashanovich, G. Z.; Nedeljkovic, M. Roadmap on silicon photonics. *J. Opt.* **2016**, *18*, No. 073003.

(3) Halir, R.; Ortega-Moñux, A.; Benedikovic, D.; Mashanovich, G. Z.; Wangüemert-Pérez, J. G.; Schmid, J. H.; Molina-Fernández, Í.; Cheben, P. Subwavelength-Grating Metamaterial Structures for Silicon Photonic Devices. *Proc. IEEE* **2018**, *106*, 2144−2157.

(4) Cheben, P.; Halir, R.; Schmid, J. H.; Atwater, H. A.; Smith, D. R. Subwavelength integrated photonics. *Nature* **2018**, *560*, 565−572.

(5) Molesky, S.; Lin, Z.; Piggott, A. Y.; Jin, W.; Vucković, J.; Rodriguez, A. W. Inverse design in nanophotonics. *Nat. Photonics* **2018**, *12*, 659−670.

(6) Piggott, A. Y.; Lu, J.; Lagoudakis, K. G.; Petykiewicz, J.; Babinec, T. M.; Vučković, J. Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer. *Nat. Photonics* **2015**, *9*, 374−377.

(7) Piggott, A. Y.; Ma, E. Y.; Su, L.; Ahn, G. H.; Sapra, N.; Vercruysse, D.; Netherton, A. M.; Khope, A. S. P.; Bowers, J. E.; Vučković, J. Inverse-Designed Photonics for Semiconductor Foundries. *ACS Photonics* **2020**, *7*, 569−575.

(8) Pond, J.; Cone, C.; Chrostowski, L.; Klein, J.; Flueckiger, J.; Liu, A.; McGuire, D.; Wang, X. In *Silicon Photonics and Photonic Integrated Circuits IV*, A complete design flow for silicon photonics; SPIE Photonics Europe: Brussels, Belgium, 2014.

(9) Hammond, A. M.; Oskooi, A.; Johnson, S. G.; Ralph, S. E. Photonic topology optimization with semiconductor-foundry design-rule constraints. *Opt. Express* **2021**, *29*, 23916−23938.

(10) Schevenels, M.; Lazarov, B. S.; Sigmund, O. Robust topology optimization accounting for spatially varying manufacturing errors. *Comput. Methods Appl. Mech. Eng.* **2011**, *200*, 3613−3627.

(11) Cabrini, S.; Kawata, S. *Nanofabrication Handbook*; CRC Press, 2012.

(12) Chrostowski, L.; Hochberg, M. *Silicon Photonics Design: from Devices to Systems*; Cambridge University Press, 2015.

(13) Owen, G.; Rissman, P. Proximity effect correction for electron beam lithography by equalization of background dose. *J. Appl. Phys.* **1983**, *54*, 3573−3581.

(14) Sentaurus Lithography, 2021https://www.synopsys.com/silicon/mask-synthesis/sentaurus-lithography.html. (accessed December 07, 2021).

(15) Pegasus Layout Pattern Analyzer, 2021. https://www.cadence.com/en_US/home/tools/digital-design-and-signoff/silicon-signoff/layout-pattern-analyzer.html. (accessed December 16, 2021).

(16) Xu, D.-X.; Schmid, J. H.; Reed, G. T.; Mashanovich, G. Z.; Thomson, D. J.; Nedeljkovic, M.; Chen, X.; Van Thourhout, D.; Keyvaninia, S.; Selvaraja, S. K. Silicon Photonic Integration Platform—Have We Found the Sweet Spot? *IEEE J. Sel. Top. Quantum Electron.* **2014**, *20*, 189−205.

(17) Lin, S.; Hammood, M.; Yun, H.; Luan, E.; Jaeger, N. A. F.; Chrostowski, L. Computational Lithography for Silicon Photonics Design. *IEEE J. Sel. Top. Quantum Electron.* **2020**, *26*, 1−8.

(18) PreFab GitHub, 2021. https://github.com/Dusandinho/PreFab. (accessed December 02, 2021).

(19) Applied Nanotools Inc. Home Page, 2021. https://www.appliednt.com/. (accessed December 02,2021).

(20) Mikołajczyk, A.; Grochowski, M. In *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, Data augmentation for improving deep learning in image classification problem; IEEE: Poland, 2018.

(21) Wong, S. C.; Gatt, A.; Stamatescu, V.; McDonnell, M. D. In *2016 International Conference on Digital Image Computing: Techniques*

and Applications (DICTA), Understanding Data Augmentation for Classification: When to Warp?; IEEE, 2016.
(22) TensorFlow Home Page, 2021. https://tensorflow.org/. (accessed February 09, 2021).
(23) Shin, H.-C.; Roth, H. R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R. M. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285−1298.
(24) Alzubaidi, L.; Zhang, J.; Humaidi, A. J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M. A.; Al-Amidie, M.; Farhan, L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, No. 53.
(25) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. 2014, arXiv:1412.6980. arXiv.org e-Print archive. https://doi.org/10.48550/arXiv.1412.6980. (accessed April 19, 2022).
(26) Sagi, O.; Rokach, L. Ensemble learning: A survey. *WIREs Data Min. Knowl. Discovery* **2018**, *8*, No. e1249.
(27) Polikar, R. Ensemble learning. In *Ensemble Machine Learning*; Springer, 2012; pp 1−34.
(28) Huang, B.; Reichman, D.; Collins, L. M.; Bradbury, K.; Malof, J. M. Tiling and stitching segmentation output for remote sensing: basic challenges and recommendations. 2018, arXiv:1805.12219. arXiv.org e-Print archive. https://doi.org/10.48550/arXiv.1805.12219. (accessed April 19, 2022).
(29) Dezfouli, M. K.; Grinberg, Y.; Melati, D.; Cheben, P.; Schmid, J. H.; Sánchez-Postigo, A.; Artega-Mońux, A.; Wangüemert-Pérez, G.; Cheriton, R.; Janz, S.; Xu, D.-X. Perfectly vertical surface grating couplers using subwavelength engineering for increased feature sizes. *Opt. Lett.* **2020**, *45*, 3701−3704.
(30) Zhang, G.; Xu, D.-X.; Grinberg, Y.; Liboiron-Ladouceur, O. Topological inverse design of nanophotonic devices with energy constraint. *Opt. Express* **2021**, *29*, 12681−12695.

https://doi.org/10.1021/acsphotonics.1c01973
ACS Photonics XXXX, XXX, XXX−XXX