SCIENCE CHINA
Information Sciences

• LETTER •

December 2023, Vol. 66 229406:1–229406:2
https://doi.org/10.1007/s11432-022-3755-y

Litho-AsymVnet: super-resolution lithography modeling with an asymmetric V-net architecture

Qing ZHANG[1†], Yuhang ZHANG[1†], Wei LU[1], Huajie HUANG[1], Zheng ZHONG[2], Congshu ZHOU[2] & Yongfu LI[1*]

[1]*Department of Micro-Nano Electronics and MoE Key Lab of Artificial Intelligence, Shanghai Jiao Tong University, Shanghai* 200240*, China;*
[2]*Primarius Technologies Co., Ltd., Shanghai* 201306*, China*

Received 25 November 2022/Revised 7 February 2023/Accepted 5 May 2023/Published online 2 November 2023

Citation Zhang Q, Zhang Y H, Lu W, et al. Litho-AsymVnet: super-resolution lithography modeling with an asymmetric V-net architecture. Sci China Inf Sci, 2023, 66(12): 229406, https://doi.org/10.1007/s11432-022-3755-y

Lithography simulation is key to the preparation of mask data and verification of mask patterns [1], which enhances design-to-wafer fidelity and minimizes distortions. However, the complexity of lithography simulation has tremendously increased with the decrease in feature size, prolonging the simulation cycle. Hence, an accurate and fast lithography simulation is in great demand.

Recently, the introduction of the machine learning method brings a new paradigm for achieving high modeling efficacy and accuracy in lithography simulation [2]. Ref. [2] has proposed LithoGAN to perform end-to-end simulation for the contact shapes, which has achieved over 96% prediction accuracy and orders of magnitude runtime reduction compared with prior studies. However, there is still room to further improve the model efficacy and speed up the simulation cycle.

In this study, we have proposed an end-to-end super-resolution lithography model, Litho-AsymVnet framework, which uses asymmetric autoencoder neural network architecture to convert the low-resolution mask patterns to high-resolution resist patterns by 6×. To further eliminate the boundary errors, we have proposed the "trimming" method and concentric weighted binary cross-entropy loss function to achieve a good trade-off between maximizing the accuracy and minimizing the runtime. The experimental results show that our Litho-AsymVnet framework achieves up to 7.5% accuracy improvement on contact/via and 2D polysilicon/metal datasets compared with the state-of-the-art studies.

Litho-AsymVnet framework. Figure 1 presents the Litho-AsymVnet framework, including the Litho-AsymVnet model and the optimization methods ("trimming" method and concentric weighted binary cross-entropy (CWBCE) loss function).

Litho-AsymVnet model architecture includes two parts: (i) encoder, and (ii) decoder, as shown in Figure 1(a). Considering the trade-off between the accuracy and speed, the Litho-AsymVnet model takes the mask pattern images with a lower resolution of $224 \times 224 \times 1$ pixel as input and outputs resist pattern images with a higher resolution of $1334 \times 1334 \times 1$. (i) The encoder network consists of four encoder blocks and a bridge block. Each encoder block consists of two 3×3 convolutional layers followed by a rectified linear unit (ReLU) layer, a batch normalization (BN) layer, and a max-pooling layer. Two 3×3 convolutional layers are used as a bridge block to connect the encoder with the decoder. The number of filters is set as 32 in the first encoder block, and it is doubled at the end of each encoder block. The max-pooling layer with a kernel size of 2×2 halves the output size. The input tensor has been reduced from $224 \times 224 \times 1$ to $14 \times 14 \times 256$ after traversing through the encoder network. (ii) The decoder network consists of five decoder blocks, which are asymmetrical as compared with the encoder network. The first four decoder blocks are composed of an up-sampling layer with the stride of 2, two 3×3 convolutional layers, a ReLU layer, and a BN layer. At the end of each decoder block, the dimensions of feature maps are doubled and the number of filters is halved. The last decoder block is similar to the first four decoder blocks while the stride of the up-sampling layer is 6, which expands the resolution of the feature maps from 224 to 1344. In the end, we adopt a 1×1 convolutional layer with a filter number of 1, followed by a sigmoid layer to generate the output. We employ skip connections to fully utilize the features from different layers and resolutions for better simulation performance [3].

Optimization methods are necessary as the neural network models do not guarantee every pixel is 100% error-free, especially near the boundary of the image [3]. In this study, we propose the "trimming" method and CWBCE loss function to achieve a good trade-off between maximizing model accuracy and minimizing the runtime. Figure 1(b) presents

* Corresponding author (email: yongfu.li@sjtu.edu.cn)
† Zhang Q and Zhang Y H have contributed equally to this work.

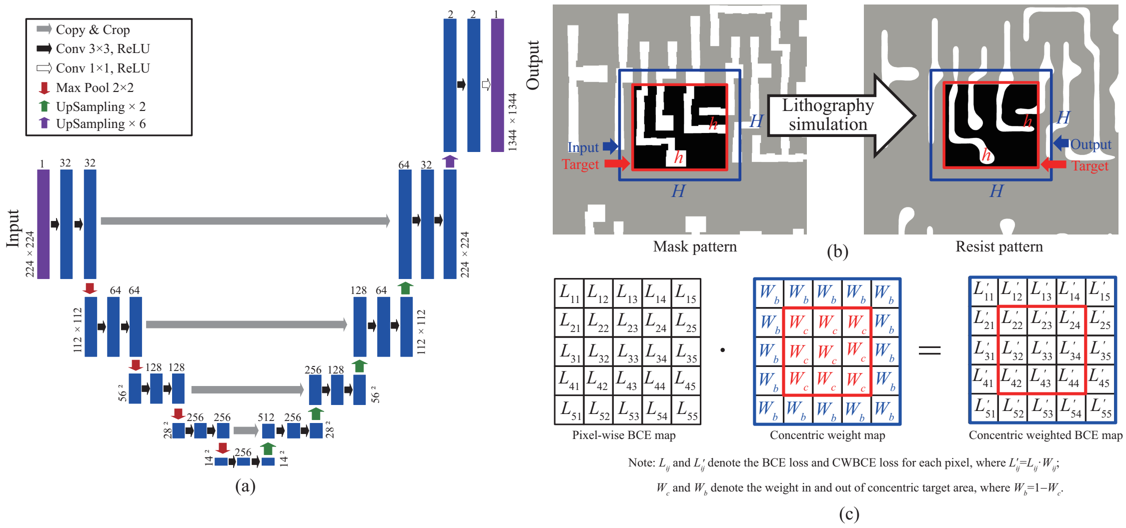

Figure 1 (Color online) The proposed Litho-AsymVnet framework, including (a) Litho-AsymVnet model architecture, (b) an illustration of trimming method, and (c) an illustration of concentric weighted binary cross-entropy (CWBCE) loss.

an illustration of the "trimming" method where the dimension of the output resist pattern image with a dimension of $H \times H$ (blue box) is trimmed to a target dimension of $h \times h$ (red box). As illustrated in Figure 1(c), the CWBCE loss function is a weighted pixel-wise loss function, where the matrix dot product is applied onto the pixel-wise BCE map with the weighted concentric map to exacerbate the error within the target area. It is formulated as follows:

$$\begin{aligned}
\mathcal{L}_{\text{cwbce}} &= \sum_{i=1}^{H}\sum_{j=1}^{H}\mathcal{L}'_{i,j} && (1)\\
&= \sum_{i=1}^{H}\sum_{j=1}^{H}\mathcal{L}_{i,j}\cdot w_{i,j} && (2)\\
&= -\sum_{i=1}^{H}\sum_{j=1}^{H}\{(1-y_{i,j})\log(1-\hat{y}_{i,j}) && (3)\\
&\quad + y_{i,j}\log(\hat{y}_{i,j})\}\cdot w_{i,j}, && (4)
\end{aligned}$$

where $\mathcal{L}'_{i,j}$ and $\mathcal{L}_{i,j}$ denote the CWBCE loss and BCE loss for each pixel. $w_{i,j}$ denotes the weight for each pixel in the weighted concentric map and $y_{i,j}$ and $\hat{y}_{i,j}$ denote the ground truth and the prediction result for each pixel, respectively. The weight of the pixel in and out of the target area is set as w_c and $1 - w_c$, respectively.

Experimental results. Our proposed Litho-AsymVnet framework is implemented and evaluated in Python on a Linux machine with a 2.50 GHz CPU and Nvidia GTX 2080Ti GPU. We quantify and evaluate the lithography model's performance using pixel accuracy, class accuracy, and mean IOU [2]. We have used 2D polysilicon and contact layers from commercial mixed-signal designs based on 55/65 nm technology nodes as our datasets. To improve the model's generalization ability, data augmentation techniques [4] are applied in this study.

We have explored the optimal target dimensions (h) to determine the best trade-off between accuracy and speed. Besides, we have conducted experiments to benchmark with the popular network architectures based on the same dataset to verify the effectiveness of our framework. Furthermore, we have conducted a comparison with LithoGAN [2] using the same datasets. The experimental results show that our framework outperforms others with the highest simulation accuracy. Detailed experimental results can be found in Appendix A.

Conclusion. In this study, we have proposed a Litho-AsymVnet framework to perform end-to-end super-resolution lithography modeling. Our Litho-AsymVnet framework with an asymmetric autoencoder architecture takes in a lower resolution mask pattern image as input and produces a 6× higher resolution resist pattern image as output. To address the boundary pixel errors, we have proposed a "trimming" method and concentric binary cross-entropy loss function to achieve a good trade-off between prediction accuracy and runtime. The experimental results show that our proposed framework produces a high quality prediction of resist pattern compared with the prior work.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 62141414, 62350610271).

Supporting information Appendix A. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Mack C A. Thirty years of lithography simulation. In: Proceedings of SPIE, 2005. 5754: 1–12
2 Ye W, Alawieh M B, Lin Y B, et al. LithoGAN: end-to-end lithography modeling with generative adversarial networks. In: Proceedings of ACM/IEEE Design Automation Conference (DAC), 2019. 1–6
3 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015. 234–241
4 Zhang Q, Zhang Y, Li J, et al. Litho-NeuralODE 2.0: improving hotspot detection accuracy with advanced data augmentation, DCT-based features, and neural ordinary differential equations. Integration, 2022, 85: 10–19