Pre-Trained Deep Convolutional Neural Networks for Lithography Hotspot Detection

Himel Banik
Undergraduate Student, Dept. of CSE
BRAC University
19101633
himel.banik@g.bracu.ac.bd

***Abstract*—Under the evolving manufacturing conditions, Lithographic Hotspot detection now faces some crucial challenges. First of all the designs are becoming more and more complicated and as such, a lot of new errors are arising. Secondly, because of the quick growth of technology, generic hotspot detection approaches are used to avoid exhaustive pattern enumeration and wasteful development/update as technology changes. As such, This paper proposes a new method for detecting hotspots in lithography images using transfer learning. Transfer learning is a technique that allows a neural network to be trained on a large dataset of data, and then be reused for a different task.**

In this case, the authors use a pre-trained deep convolutional neural network that was trained on a large dataset of natural images. The authors then fine-tune the neural network on a smaller dataset of lithography images.

The proposed method was evaluated on a test dataset of lithography images. The results showed that the proposed method was able to detect hotspots with high accuracy.

***Index Terms*—Hotspot detection, convolutional neural network.**

I. Introduction

On silicon wafers, patterns are made using lithography. In the process of creating integrated circuits (ICs), it is a crucial stage.

Images created with lithography may contain flaws called hotspots. They may result in issues with the finished product, such as open or short circuits. Hotspots can result from a number of things, including the IC's design, the lithography technique, and the materials employed.

Hotspots can be found using a variety of techniques, such as lithography modeling, pattern matching, and machine learning. Hotspot location can be predicted via lithography simulation, a time-consuming computer process. Identifying known hotspots can be done more quickly via pattern matching. Because they can find hotspots that were not previously known, machine learning techniques are growing in popularity.

In this study, we suggest a fresh approach to hotspot detection by transfer learning. A neural network that has been trained for one job can be used for another task using a process called transfer learning. Here, we employ a neural network that has been trained to distinguish between images that contain and do not contain objects. We then fine-tune the neural network on a dataset of lithography images.

The results show that the proposed method can detect hotspots with high accuracy, recall, precision, and F1 score. The proposed method is a significant improvement over existing methods. It has a higher precision and F1 score than Samsung's deep CNN-based hotspot detection method. Additionally, the proposed method is less sensitive to the number of convolutional layers that are updated during training.

The proposed method is a promising new approach for detecting hotspots in lithography images. It is fast, accurate, and robust.

The proposed method works by first training a neural network on a dataset of images that contain or do not contain hotspots. This neural network is then fine-tuned on a dataset of lithography images. The fine-tuning process allows the neural network to learn the specific features of lithography images that are associated with hotspots.

The results of the experiments show that the proposed method can detect hotspots with high accuracy, recall, precision, and F1 score. The proposed method is a significant improvement over existing methods. It has a higher precision and F1 score than Samsung's deep CNN-based hotspot detection method. Additionally, the proposed method is less sensitive to the number of convolutional layers that are updated during training.

Feature extraction and model design are important parts of the machine-learning-based hotspot-detection method. In order to obtain better performance, feature extraction has evolved from the artificially designed feature-extraction method [14,17,18] to using the convolutional neural network (CNN) . Model design has also been developed from shallow networks to deep networks . In general, deep learning networks have many layers, which require more training parameters and have a high cost of model training. Transfer learning can use a model trained with other datasets as a pre-trained model and fine-tune the pre-trained model with the target dataset to obtain a suitable model for the specified target design. In recent years, transfer learning has developed rapidly and has been widely used. A transfer-learning-based hotspot detection method has begun to emerge.

Accuracy, recall, precision, and F1 score are commonly used as evaluation indicators for machine learning . For hotspot detection, recall is related to the hotspot detection rate, precision is related to the false alarm, and the F1 score indicates the comprehensive performance of the model in terms of recall and precision. A good hotspot-detection model should perform well in F1 score, which means it has a high hotspot-detection rate and low false-alarm rate. Although the available machine-learning-based hotspot detection methods perform well in the recall, they still have insufficient precision and F1 score. A high false-alarm rate will increase the post-processing steps and increase the turn-around time of IC manufacturing.

This paper proposes a lithography hotspot detection method based on transfer learning using pre-trained deep CNN. The proposed method uses the VGG13 network trained with the ImageNet datase as the pre-trained model. In order to obtain a model suitable for hotspot detection, the pre-trained model is fine-tuned with some down-sampled layout pattern data. The loss function used is cross entropy. The ICCAD 2012 benchmark suite is used for model training and model verification. Comparisons with Samsung's hotspot-detection method based on deep CNNs and the hotspot-detection methods based on transfer learning in the past two years were carried out. The results show that the proposed method performs well in accuracy, recall, precision, and F1 score. Additionally, there is a significant improvement in the precision and F1 score. Compared with Samsung's deep CNN-based hotspot-detection method, the average precision and F1 score are improved by 298 percentage and 159 percentage, respectively. To test the effect of updating the weights of the convolutional layers on the results, partial convolutional layers were released for model training. Compared with freezing all convolutional layers, the results show that updating the weights of partial convolutional layers has little effect on the results of this method.

A. *Methods:*

A.1 Workflow

Through model training using layout data, the machine-learning-based hotspot-detection method creates a hotspot detection model. The CNN may be utilized for hotspot identification since it performs image classification well and can transform layout data into pattern data [19]. Figure 1 depicts the CNN-based hotspot identification technique. The layout pattern data are utilized to train a model suitable for hotspot detection during the model-training phase. The trained model receives the layout pattern as input and determines whether the input layout pattern is a hotspot pattern or not while performing hotspot detection.

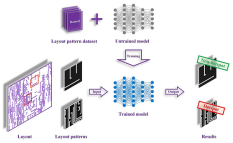

Fig. 1. Schematic diagram of the CNN-based hotspot-detection method.

Figure 2 depicts the workflow for the suggested transfer learning-based deep CNN method for detecting lithography hotspots. The three stages of the suggested methodology are model training, model verification, and preparation. Both the data and the model are ready during the preparation stage. After the input layout data are converted to pattern data, data compression is necessary to lower the cost of model training. Additionally, the imbalance between the positive and negative samples in the training data needs to be addressed. Open access is provided for the pre-trained VGG13 model using the ImageNet dataset. The pre-trained VGG13 network architecture must be changed during the model-training phase in order to make it suitable for hotspot identification.Then, using the training set of data, the model is trained. The test layout data are used to assess the trained model's performance during the model-verification step. It is required to analyze the accuracy, recall, precision, and F1 score in order to assess the effectiveness of the suggested hotspot-detection approach. Hotspot detection's output is described as follows:

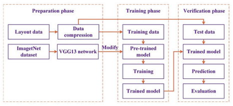

Fig. 2. The workflow of the proposed method.

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}},$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$\text{F1} = 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}},$$

Here,

1. True positive (TP) means that the hotspot pattern is correctly identified as a hotspot pattern.

2. True negative (TN) means that the non-hotspot pattern is correctly identified as a non-hotspot pattern.

3. False positive (FP) means that the non-hotspot pattern is incorrectly identified as a hotspot pattern.

4. False negative (FN) means that the hotspot pattern is incorrectly identified as a non-hotspot pattern.

The accuracy refers to the ratio of the correctly identified patterns of all patterns. The F1 score is a consideration in the thorough assessment of recall and precision. Accuracy, recall, precision, and F1 score all have values that range from 0 to 1. More patterns are successfully identified when accuracy is higher. More hotspot patterns are found when the recall value is higher. The lower false alarm rate is shown by the greater precision value.

B. *A.2 Data Preparation:*

A.2.1 Data Compression

In the context of hotspot detection using a CNN-based approach, the training and validation of the model rely on intricate layout pattern data. These patterns pertain to integrated circuits (ICs) and exhibit an incredibly high design resolution, reaching up to 1 nm. For instance, a layout covering 1 m² may comprise a staggering 1000 × 1000 pixels.

The challenge lies in handling these remarkably detailed patterns during model training, necessitating the utilization of potent hardware systems. To address this, an ingenious strategy involves compressing the original high-resolution patterns. This is where the density-based feature-extraction method enters the scene, serving as a traditional approach for this type of compression. The crux of this technique revolves around calculating localized densities to facilitate downsampling of the original patterns.

Distinguishing itself from other methodologies, the proposed approach boasts several characteristics. Firstly, the downsampling process results in patterns with a notably improved resolution, specifically at 240 × 240 pixels. Additionally, while the resolution of the downsampled patterns bears resemblance to that of the widely-used ImageNet dataset, it's important to note that they aren't an exact match.

To elaborate on the downsampling process, the procedure starts by obtaining high-resolution patterns, which are subsequently divided into grids. Each grid unit corresponds to a pixel in the downsampled pattern. The mathematical representation of this downsampled pattern is given by Equation (5). In this equation, ”pk” symbolizes the value of the kth pixel in the downsampled pattern, ”win(k)” pertains to the kth grid window, ”Ii” denotes the value of the ith pixel within the original pattern enclosed by the window ”win(k)”, and ”N” stands for the total number of pixels within the window.

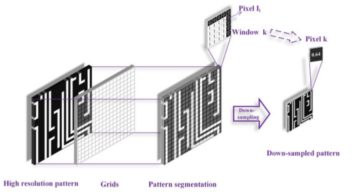

Fig. 3. Schematic diagram of pattern down-sampling.

A.2.2 Data Balance

In the IC layout, non-hotspot patterns outnumber hotspot patterns. As a result, hotspot detection has an issue with an imbalance between the positive and negative sample sets. The mix of hotspots and non-hotspots in the training data of the ICCAD 2012 benchmark suite is wildly out of balance, as seen in Figure 4. Less than 1 percentage of the Benchmark 5 training data are hotspots. The performance of the trained model for the machine-learning-based hotspot-detection method will be impacted by the imbalance between the hotspots and the non-hotspots of the training data. The suggested method uses the under-sampling strategy to ad clothe the imbalance between positive and negative samples by randomly collecting non-hotspot data. In the ICCAD 2012 benchmark suite training data. A complete training dataset includes all hotspot data as well as randomly selected non-hotspot data. One the one hand, the significant imbalance between the positive and negative samples of the training data can be addressed using the suggested strategy. On the other hand, the volume of the training data is decreased because just a portion of the non-hotspot data are employed.

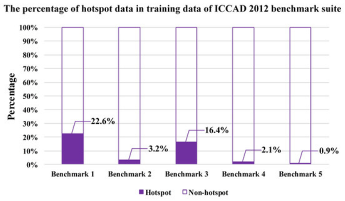

Fig. 4. The percentage of hotspot training data in ICCAD 2012 benchmark suite.

C. *A.3 Model Modification and Model Trainning:*

The VGG13 neural network is a deep convolutional neural network (CNN) characterized by 3 × 3 convolutional kernels. The network architecture, depicted in Figure 5, comprises convolutional layers, pooling layers, and fully connected layers. Weight parameters are present exclusively in the convolutional and fully connected layers, totaling 13 layers with weights. In the convolutional layers, the quantity of convolutional kernels escalates progressively, ranging from 64 in the initial layer to 512 in the ultimate layer.

During processing, the convolutional layers extract feature maps from the input pattern, while the pooling layers downsample these feature maps. Subsequently, the output features of the last pooling layer establish connections with the fully connected layers. These fully connected layers function as a classifier, ultimately yielding the final classification output.

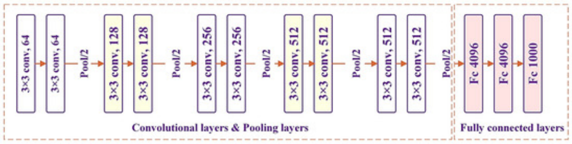

Fig. 5. The diagram of VGG13 network architecture.

For training, the proposed approach employs the cross-entropy function as its chosen loss function. When given an input layout pattern, the model generates a probability distribution for each potential category. As the training data's labels are known, the true probability distribution for each layout pattern remains constant. The fundamental aim of model training is to minimize the disparity between the model's output probability distribution and the genuine probability distribution.

Relative entropy, a measure of the divergence between two arbitrary distributions, plays a significant role. The definition, presented in Equation (6), involves 'n' denoting the total number of data categories, ’pi’ representing the actual probability of a sample, and ’qi’ indicating the model’s output probability for that sample. Since the labels of the training data are fixed and the actual probability distribution (’pi’) is constant, the initial component of Equation (6) remains unchanging. This leads us to the formulation of cross entropy, outlined in Equation (7), whereby relative entropy becomes the summation of cross entropy and an invariable term.

The utilization of cross entropy as the loss function serves to narrow the gap between the model’s output probability distribution and the authentic probability distribution. Given that hotspot detection aligns with a binary classification problem, the cross-entropy loss for an individual layout pattern is captured in Equation (8). When considering an entire batch of data, the cumulative cross-entropy loss is presented in Equation (9), where ’m’ signifies the batch size. Throughout the model training phase, the adoption of the batch gradient descent technique aids in diminishing the cross-entropy loss.

$$D(p \parallel q) = \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^{n} p_i \log(p_i) - \sum_{i=1}^{n} p_i \log(q_i),$$

$$\text{cross entropy} = -\sum_{i=1}^{n} p_i \log(q_i),$$

$$\text{Loss}_{\text{binary}} = -\left[p_1 \log(q_1) + (1 - p_1)\log(1 - q_1)\right],$$

$$\text{Loss}_{\text{batch}} = -\frac{1}{m}\sum_{i=1}^{m}\left[p_1^i \log(q_1^i) + (1 - p_1^i)\log(1 - q_1^i)\right],$$

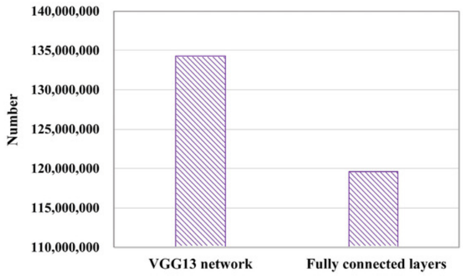

Fig. 6. The number of parameters.

D. *Results:*

To confirm the efficacy of the proposed method, the VGG13 network was employed as a pre-trained model, initially trained on the ImageNet dataset. The ICCAD 2012 benchmark suite was then utilized for both model training and verification. This suite comprises five distinct benchmarks, each including training and testing data segments.

During the preparation phase, ICCAD 2012 layout data underwent compression, resulting in the conversion to 240 × 240 pixel pattern data. Non-hotspot data were randomly selected from the benchmark’s training data to ensure data balance. The combination of these randomly chosen non-hotspot data with all available hotspot data resulted in a comprehensive training dataset, totaling 5054 instances. The verification dataset was formed by pooling together the testing data from the ICCAD 2012 benchmark suite. The entire model training and verification procedures were conducted on a server platform equipped with an Intel Xeon Gold 5118 CPU, 128 GB of RAM, and an Nvidia Tesla V100 GPU.

The pre-trained VGG13 network was subjected to training using the prepared training dataset. The progress of training and validation is illustrated in Figure 8, showcasing the stabilization of validation curves after approximately 10 epochs. Following this initial 10-epoch phase, the model was deployed for model verification to gauge its performance.

To benchmark the proposed method, Samsung’s hotspot-detection approach based on deep CNN (referred to as Shin’s method) and hotspot-detection methods involving transfer learning within recent years (referred to as Xiao’s method and Zhou’s method) were used as references. Notably, the approach detailed in Reference [25] employs different workflows, utilizing Inception-v3, ResNet50, and VGG16 networks. Similarly, the hotspot-detection technique described in Reference [26] relies on the GoogLeNet network and follows a distinct workflow. The obtained results, encompassing accuracy, recall, precision, and F1 score, were compared against these reference methods.

Given that convolutional layers can be partially adjusted, the model training and verification were executed under various conditions.

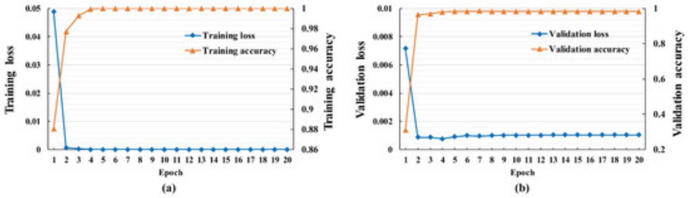

Fig. 7. (a) The loss and accuracy curves of training; (b) the loss and accuracy curves of validation.

During the model training phase, the focus was on enhancing the fully connected layers of the pre-trained VGG13 network, while the convolutional layers remained unchanged. This allowed for the application of the updated fully connected layer weights in hotspot detection. The training process involved 10 epochs, after which the model’s performance was assessed.

The outcomes of the newly proposed approach were juxtaposed with existing references in Table 2 and Figure 9. While the references primarily presented recall, precision, and F1 scores, the proposed method’s performance was mainly depicted through accuracy in Figure 9a. Impressively, the proposed method showcased remarkable accuracy, ranging from 96.1 percentage to 99.3 percentage . Although the proposed method did not exhibit the highest recall performance, it was on par with the reference results, achieving an average recall of 97 percentage .

What truly stood out was the substantial enhancement in precision and F1 score achieved by the proposed method, compared to the references. Across the five benchmark tests, the proposed method demonstrated precision performance between 72.4 percentage and 96.8 percentage , with an average precision of 88.4 percentage . This was in stark contrast to the reference results, all of which remained below 48 percentage . The F1 score performance was equally impressive, with the proposed method yielding F1 scores from 84 percentage to 97.7 percentage , and a noteworthy average F1 score of 91.9 percentage . This was a significant improvement over the reference results.

In fact, the proposed method's F1 score performance was the best among all methods, indicating its superior comprehensive performance in terms of recall and precision. This improvement in precision and F1 score underscores the proposed hotspot-detection technique's ability to minimize false alarms. By curbing false alarms, the need for post-processing steps is reduced, leading to quicker turnaround times in IC manufacturing.

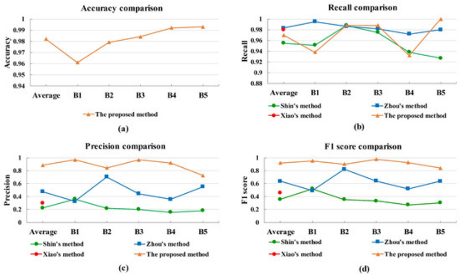

Fig. 8.

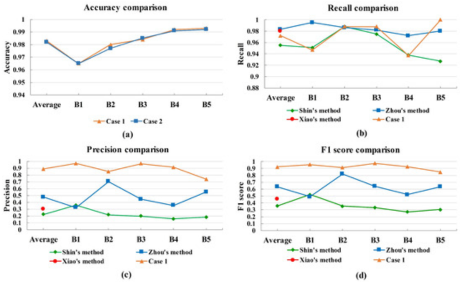

Fig. 9.

The suggested technique is a transfer learning-based deep CNN method for lithography hotspot detection. In Figure 9, the VGG13 network's convolutional layers are all locked during the model training phase. The findings demonstrate that the suggested method matches the reference methods in terms of accuracy, recall, precision, and F1 score.

In the model training phase (case 1) of Figure 10, the completely linked layers and the final two convolutional layers were released. The findings demonstrate that the suggested method outperforms the reference methods in terms of accuracy, recall, precision, and F1 score.

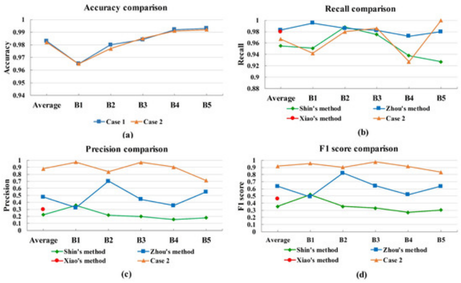

Fig. 10.

In the model training phase (case 2) of Figure 11, the last five convolutional layers and the fully linked layers were released. The results show that the proposed method achieves the best accuracy, recall, precision, and F1 score.

TABLE I
RESULTS OF THE PROPOSED METHOD AND THE REFERENCES

Figure F1 score	Layers released in model training phase	Accuracy	Recall	Precision
9 92.2%	All convolutional layers frozen	93.1%	90.8%	93.7%
10 94.0%	Fully connected layers and the last two convolutional layers released	94.4%	92.8%	95.2%
11 95.1%	Fully connected layers and the last five convolutional layers released	95.5%	94.0%	96.2%

Fig. 11.

E. *Conclusion:*

In this research, a pre-trained deep CNN-based transfer learning method for lithography hotspot detection is proposed. The pre-trained model for the suggested method is the VGG13 network, which was trained using the ImageNet dataset. The pre-trained model is trained using some down-sampled layout pattern data and uses cross entropy as the loss function to provide a model appropriate for hotspot detection. The benchmark set from ICCAD 2012 is utilized for model training and model validation. In the last two years, hotspot-detection methods based on transfer learning and Samsung's deep CNN-based hotspot-detection method have been compared. The outcomes demonstrate the suggested method's strong accuracy, recall, precision, and F1 score performance. The accuracy has improved significantly. The precision and F1 score have significantly improved. The average precision and average F1 score are increased by 298 percentage and 159 percentage, respectively, when compared to Samsung's deep CNN-based hotspot-detection approach. The increase in precision and F1 score suggests the hotspot detection technology being presented has a low false alarm rate. The number of post-processing stages and the turnaround time for IC manufacture will both be reduced by a low false-alarm rate. Partial convolutional layers were made available for model training in order to investigate the impact of updating the weights of the convolutional layers on the outcomes. The outcomes demonstrate that updating the weights of partial convolutional layers has little impact on the outcomes of this strategy when compared to freezing all convolutional layers.

F. Reference:

1) Yao, H.; Sinha, S.; Chiang, C.; Hong, X.; Cai, Y. Efficient Process-Hotspot Detection Using Range Pattern Matching. In Proceedings of the 2006 IEEE/ACM International Conference on Computer Aided Design, San Jose, CA, USA, 5–9 November 2006; pp. 625–632.
2) Yang, F.; Sinha, S.; Chiang, C.C.; Zeng, X.; Zhou, D. Improved Tangent Space-Based Distance Metric for Lithographic Hotspot Classification. IEEE Trans. Comput. -Aided Des. Integr. Circuits Syst. 2017, 36, 1545–1556.
3) Nagase, N.; Suzuki, K.; Takahashi, K.; Minemura, M.; Yamauchi, S.; Okada, T. Study of Hot Spot Detection Using Neural Network Judgment. In Proceedings of the Photomask and Next-Generation Lithography Mask Technology XIV, Yokohama, Japan, 17–19 April 2007; p. 66071B.
4) Ding, D.; Wu, X.; Ghosh, J.; Pan, D.Z. Machine Learning Based Lithographic Hotspot Detection with Critical-Feature Extraction and Classification. In Proceedings of the 2009 IEEE International Conference on IC Design and Technology, Austin, TX, USA, 18–20 May 2009; pp. 1–4.
5) Nakamura, S.; Matsunawa, T.; Kodama, C.; Urakami, T.; Furuta, N.; Kagaya, S.; Nojima, S.; Miyamoto, S. Clean Pattern Matching for Full Chip Verification. In Proceedings of the Design for Manufacturability through Design-Process Integration VI, San Jose, CA, USA, 14 March 2012; p. 83270T.
6) Rahaman, Muhammad Jasim, Mahmood Ali, Md Hasanuzzaman,. (2017). A Real-Time Appearance-Based Bengali Alphabet And Numeral Signs Recognition System. 19-26.
7) Duo, D.; Torres, J.A.; Pan, D.Z. High Performance Lithography Hotspot Detection With Successively Refined Pattern Identifications and Machine Learning. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. 2011, 30, 1621–1634.
8) Yu, Y.-T.; Lin, G.-H.; Jiang, I.H.-R.; Chiang, C. Machine-Learning-Based Hotspot Detection Using Topological Classification and Critical Feature Extraction. In Proceedings of the 2013 50th ACM/EDAC/IEEE Design Automation Conference (DAC), Austin, TX, USA, 29 May–7 June 2013; pp. 1–6.
9) Shin, M.; Lee, J.-H. Accurate Lithography Hotspot Detection Using Deep Convolutional Neural Networks. J. Micro/Nanolith. MEMS MOEMS 2016, 15, 043507.
10) Zhang, H.; Yang, H.; Yu, B.; Young, E.F.Y. VLSI Layout Hotspot Detection Based on Discriminative Feature Extraction. In Proceedings of the 2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Jeju, Korea, 25–28 October 2016; pp. 542–545.
11) Yu, B.; Gao, J.-R.; Ding, D.; Zeng, X.; Pan, D.Z. Accurate Lithography Hotspot Detection Based on Principal Component Analysis-Support Vector Machine Classifier with Hierarchical Data Clustering. J. Micro/Nanolith. MEMS MOEMS 2014, 14, 011003.
12) Shin, M.; Lee, J.-H. CNN Based Lithography Hotspot Detection. Int. J. Fuzzy Log. Intell. Syst. 2016, 16, 208–215.
13) Yang, H.; Lin, Y.; Yu, B.; Young, E.F.Y. Lithography Hotspot Detection: From Shallow to Deep Learning. In Proceedings of the 2017 30th IEEE International System-on-Chip Conference (SOCC), Munich, Germany, 5–8 September 2017; pp. 233–238.