# A DATA-DRIVEN APPROACH FOR AUTOMATED INTEGRATED CIRCUIT SEGMENTATION OF SCAN ELECTRON MICROSCOPY IMAGES

Zifan Yu<sup>\*</sup> Bruno Machado Trindade<sup>†</sup> Michael Green <sup>†</sup> Zhikang Zhang <sup>\*</sup> Pullela Sneha<sup>†</sup> Erfan Bank Tavakoli<sup>\*</sup> Christopher Pawlowicz<sup>†</sup> Fengbo Ren<sup>\*</sup> <sup>\*</sup> Arizona State University, Tempe, AZ, USA <sup>†</sup>TechInsights Inc., Ottawa, ON, Canada

### ABSTRACT

This paper proposes an automated data-driven integrated circuit segmentation approach of scan electron microscopy (SEM) images inspired by state-of-the-art CNN-based image perception methods. Based on the requirements derived from real industry applications, we take wire segmentation and via detection algorithms to generate integrated circuit segmentation maps from SEMs in our approach. On SEM images collected in the industrial applications, our method achieves an average of 50.71 on Electrically Significant Difference (ESD) in the wire segmentation task and 99.05% F1 score in the via detection task, which achieves about 85% and 8% improvements over the reference method, respectively.

*Index Terms*— image segmentation, deep learning, scan electron microscopy images, integrated circuit segmentation

# 1. INTRODUCTION

Integrated circuit segmentation (ICS) that extracts circuits from scan electron microscopy (SEM) images is a critical task in semiconductor analysis. The limited number of existing public ICS approaches require much human intervention to generate accurate segmentation results [1, 2]. Such human interventions are usually undesired and even unacceptable for large-scale industrial purposes, significantly limiting real-industrial applications. More specifically, [2] requires manually tuning model parameters and separation thresholds in the segmentation process based on the visual appearance of the segmentation results. Ronald et al. [1] propose a histogram-based ICS approach that uses decision boundaries derived from the peaks in the intensity histogram to perform wire segmentation. However, due to the high variation of intensity values in our real-industry collected SEM images, this approach fails to generate highly accurate segmentation results. Additionally, [1, 2] focus on wire segmentation without exporting the location information of vias (i.e., electrical connections between copper layers in ICs) which are highly demanded in semiconductor analysis. The deep learning-based data-driven approach [3, 4, 5] suffers from high segmentation error rates caused by random and intensive noise and contamination on SEM images we captured in the industry. Such errors in the segmentation results require laborious manual corrections by experts afterward.

In this paper, we propose a data-driven ICS approach that automatically generates circuit segmentation results from SEM images inspired by the state-of-the-art CNN-based image perception methods in other domains, e.g., medical microscopy imaging [6, 7, 8]. The proposed approach runs in three steps: pre-processing, image segmentation, and postprocessing (see Figure 1). The pre-processing step runs SEM image patch generation and on-the-fly training data augmentation. Then, we perform CNN-based wire segmentation and via detection separately to derive wire and via integrated segmentation maps in image segmentation step [9, 10, 6], where CNN networks are modified to fit our domain-specific SEM images. In terms of post-processing, we propose a pixel-wise refiner on the coarse segmentation maps to reduce electrical shorts and opens caused by artifacts containing small areas of isolated pixels. We also merge the vias detected inside the overlapping area between patches and drop the vias near patch edges. On SEM images of 7 different types of ICs (e.g., microprocessor and power management ICs) we collected in industrial applications, our approach achieves an average of 50.71 on Electrically Significant Difference (ESD) [2] in the wire segmentation task and 99.05% F1 score in the via detection task, which achieves about 85% and 8% improvements over the reference method [3], respectively.

# 2. METHODOLOGY

Our proposed approach consists of three major steps, as shown in Figure 1: pre-processing, image segmentation, and post-processing.

# 2.1. Pre-processing

Since our SEM images have much higher resolution (i.e.,  $8192 \times 8192$ ) than input sizes (e.g.,  $256 \times 256/512 \times 512$ ) of typical CNN-based image segmentation approaches, we perform image segmentation on smaller image patches instead of full-sized images. As smaller image patches inevitably lose the global information beneficial for image segmentation, such an approach could lead to a lower segmentation

This work is supported by a research contract from TechInsights Inc. Corresponding author email: zifanyu@asu.edu.



Fig. 1: The pipeline of our proposed approach

accuracy[6, 11]. However, we observe that most SEM images contain highly repetitive patterns at areas of wire/via in each image; the intensity difference between the background and wire areas are significant, which implies that the texture information of a local patch is sufficient for wire segmentation. We also observe that vias might be incomplete in an SEM patch due to the patch-cutting operation when vias lie exactly at the edges of such a patch. Therefore, we generate patches with a 100-pixel overlapping area for each pair of adjacent patches to ensure every via can be entirely shown at least in one image patch.

We utilize standard data augmentation methods, including vertical and horizontal flip, random intensity augmentation, and  $n \times 90$  degree rotation to enlarge the limited labeled training data. We do not utilize random rotation for data augmentation in our approach since wires and vias are always vertical or horizontal in SEM images. We randomly generate a uniformly distributed number between -5 and 5 as a jitter for each unsigned 8-bit pixel. The random intensity augmentation introduces minor noises to training samples which can improve the robustness of the trained model to noises.

#### 2.2. Image Segmentation

# 2.2.1. Wire Segmentation

As the high-level semantic features, i.e., the information carried by low-resolution feature maps, are not critical for our SEM image segmentation, we avoid the unnecessary feature map downsampling in the original HRNet[6]. We modify the stride of the second CNN layer to 1 so that our network starts extracting visual features from feature maps with  $\frac{1}{2}$  of the original input size instead of  $\frac{1}{4}$ . Then, we remove the fourth parallel CNN stage of the original HRNet[6]. Consequently, the modified network extracts and merge multi-resolution visual features at three different resolutions, which are  $\frac{1}{2}$ ,  $\frac{1}{4}$ , and  $\frac{1}{8}$  of the original input size without the  $\frac{1}{16}$  ones in the original HRNet. The output feature maps with the largest size of the last stage are upsampled to the same size as the original SEM patch and fed into the final classification layer. The final classification layer is a CNN layer with kernel = 1 and stride = 1. The classification layer outputs a binary segmentation result of the input SEM patch. The loss function for our wire segmentation model training is pixel-level binary class cross-entropy, which is

$$L_{wire}(y_{gt}, y_{pred}) = -(y_{gt}\log(y_{pred}) + (1 - y_{gt})\log(1 - y_{pred})),$$
(1)

where  $y_{gt}$  is the ground truth label and  $y_{pred}$  is the predicted label.

# 2.2.2. Via Detection

We utilize the Faster R-CNN[10] as our fundamental algorithm for via detection. We replace the ROI pooling layer in the original Faster R-CNN with the ROI align layer that samples the proposed region from feature maps more accurately using interpolation, which results in an about 8% detection performance improvement in the natural image object detection. We use HRNet[6] and ResNet[12] as visual feature extraction networks separately to provide multi-resolution feature maps for the detection head. The loss function of our via detection network is  $L_{via} = L_{rpn} + L_{box}$ , where  $L_{rpn}$ and  $L_{box}$  are the loss of the region proposal network and the bounding box regression loss in Faster R-CNN[10], respectively.

#### 2.3. Post-processing

# 2.3.1. Neighbor Pixel Refiner for Wire Segmentation

As shown in Figure 2, we notice some ESDs are caused by isolated wire pixels. Such ESDs can be eliminated by merging the isolated pixels into nearby wires or dropping the isolated pixels. Hence, we propose a wire segmentation refiner, where each pixel is re-classified according to coarse segmentation labels of its neighbor pixels. This refiner is implemented using GPU-accelerated fix-weight convolution operations. For



**Fig. 2**: Visual examples of ESDs caused by isolated areas of pixels. FP wires are pixels labeled incorrectly as wires in predictions. The ESD A and B are caused by isolated pixels in the red boxes.

a pixel P, the convolutional kernel K selects  $k^2 - 1$  neighbors around P. Elements of K are initialized as 1 except the center element. As the segmentation results contain binary classes, i.e., "1" is wire and "0" is background, the convolution output c equals the number of predicted wire pixels around P. If c is greater than the threshold t, we re-classify P as wire pixel and vice versa.

### 2.3.2. Merge-Overlapped Predictions for Via Detection

The output of the via detection is a list of predicted boxes and the corresponding confidence score for each prediction. Note only predictions with a confidence score greater than 0.6 are kept. We eliminate predicted vias totally inside a 50pixel border area of each patch. Correct vias among these dropped predictions can still be detected in other neighbor patches where they are entirely presented, but the border area can effectively reduce the number of false-positive cases. We match the overlapped predictions in the neighbor patches by computing Intersection over Union (IoU) between any pair of predicted boxes in the results. If the IoU between two predictions is greater than 0.3, we view these two predictions as one predicted via and keep the one with higher confidence scores.

#### **3. EXPERIMENTS**

### 3.1. Experiment Setup

Our dataset contains SEM images of the microprocessor, Radio Frequency (RF) transceiver, power management, flash memory, SoC, and so on with 2.92 nm average pixel size and  $22.96 \mu\text{m}$  average field size, respectively. The collecting dwelling time is  $0.2 \,\mu\text{s}/\text{pixel}$ , and our scanning is  $50 \times$  faster than data collected by [1] so that our SEM images contain more noise. We train HRNet[6] for 100 epochs with 21 highresolution SEM images for wire segmentation. We use the Adam optimization algorithm with an initial learning rate of

 
 Table 1: Wire segmentation results. Our approach outperforms the reference method by large margins.

Models	Avg ACC	Avg IoU	Avg ESD
FCN with VGG16[3]	94.40%	89.32%	329.86
HRNet-3	<b>95.75</b> %	<b>91.86</b> %	50.71
HRNet-4	95.71%	91.78%	69.77

**Table 2**: The performance improvements contributed by the neighbor pixel refiner. The neighbor pixel refiner shows consistent and significant improvements across all the parameter setups.

Models	k	t	Refiner		RB
			w/o	w/	КК
HRNet-3	7	24	66.38	55.90	15.79%
	7	36		50.71	23.61%
	9	40		55.71	16.07%
HRNet-4	7	24	82.67	69.77	15.60%
	9	40		69.86	15.50%

0.001 and a weight decay of  $10^{-8}$ . We decay the learning rate by a factor of 0.1 if the loss on the validation set stops reducing in 2 epochs. For via detection, we train the networks, i.e., Faster R-CNN [10] with HRNet[6] and ResNet[12] separately, for 150 epochs with 100 high-resolution SEM images. We use the stochastic gradient descent (SGD) optimizer with an initial learning rate of 0.001, which is decayed by a factor of 10 every 30 epochs, a momentum of 0.9, and a weight decay of  $5 \times 10^{-4}$ .

#### 3.2. Wire Segmentation

We evaluate the effectiveness of our approach in wire segmentation using pixel-level classification accuracy and IoU, and the improved ESD evaluation method proposed by [2]. ESDs are shorts or opens that do not exist in ground truth circuits and result in electrical errors, which lower value indicates better segmentation performance. However, note that wrong classified pixels may not cause shorts or opens in extracted circuits.

We present the quantitative results on 21 high-resolution results in Table 1. Compared to HRNet-4, which has the same number of parallel CNN stages as the original HRNet[6], our modified HRNet-3 achieves higher average accuracy and average IoU, which indicates the low-level texture features are enough for wire segmentation. Additionally, the more obvious average ESD gap between HRNet-3 and HRNet-4 indicates that some misclassified pixels do not cause shorts or opens in the extracted circuit segmentation. Compared with the reference method, our method reduces 85% of ESDs and improves the accuracy and IoU by 1.35% and 2.54%, respectively. Parameters k and t in Tabel 2 are defined in Section2.3.1 and reduced rate (RR) is the percentage of the ESDs reduced by the proposed refiner. When we apply a refiner with k = 7 and t = 0.5, we can reduce at least 15.6% ESDs in the coarse segmentation results generated by HRNet-3 and HRNet-4, which indicates that our refiner can be applied to coarse segmentation results automatically without hand-tuning the kernel size or threshold. Figure 3 shows the qualitative comparison of our wire segmentation method with the reference method[3].



**Fig. 3**: Visual examples of wire segmentation results. Our approach has fewer ESDs than the reference method.

**Table 3**: Via detection results. Our approach outperforms the reference method by large margins. P is precision and R is recall. FR-CNN means Faster R-CNN.

Models	Р	R	F1
FCN w/ VGG16[3]	93.44%	89.35%	91.35%
FR-CNN w/ HRNet-4	99.72%	98.38%	<b>99.05</b> %
FR-CNN w/ HRNet-5	<b>99.</b> 77 <i>%</i>	98.23%	98.99%
FR-CNN w/ ResNet	98.88%	<b>98.56</b> %	98.72%

**Table 4**: Via detection results of Faster R-CNN with HRNet-4 on overlapping or non-overlapping patches. These results indicate that higher F1 can be achieved by utilizing overlapping patches(OP) during inference.

Models	Precision	Recall	F1
w/ OP	<b>99.</b> 77 <i>%</i>	<b>98.34</b> %	99.05%
w/o OP	94.40%	94.62%	94.51%

# 3.3. Via Detection

We use the precision to evaluate the error rate in predictions and recall to evaluate the via retrival rate. F1 score[13] combins precision and recall and presents the overall performance



**Fig. 4**: Visual examples of via detection results. Red boxes are the detection results of [3]. Green boxes are predictions of our method. Our approach detects more vias correctly than the reference method not only for the locations but also for the size.

of methods. If a predicted box has IoU greater than 0.3 with any ground truth box, we refer to the prediction as a true positive case. Note that each ground truth box can only have one matched predicted box, which is the predicted box that has the largest IoU. A predicted box without matched ground truth box is a false positive case, and a ground truth box without matched predicted box is a false negative case.

We present evaluation results on 20 high-resolution SEM images in Table 3. Compared to the reference method[3], our approach achieves an 8.4%, 6.72%, and 10.11% improvement on F1 score, precision, and recall, respectively. Faster R-CNN[10] with HRNet[6] as backbone outperforms the original Faster R-CNN which utilizes the ResNet[12] as CNN backbone for visual feature extraction with an 0.33% F1 score improvement. Furthermore, We explore the impact of generating overlapping patches for via detection inference. The model inference with overlapping patches effectively achieves a 4.54% F1 score improvement, indicating that removing the predictions in border area can reduce the number of incorrectly detected via-like objects, and generating patches with an overlapping area makes the model inference more robust. We present the qualitative visualization result comparison of via detection in Figure 4. Note that the reference method [3] generates vias with irregular shapes and sizes, so we replace the vias with rectangles. In contrast, our via detection method generates vias with a regular size closer to the real vias in the SEM images.

# 4. CONCLUSION

This paper proposes an automatic data-driven approach for ICS, which is able to generate IC segmentataion maps and locations of vias in an end-to-end manner without any human intervention. Our approach achieves an average of 50.71 on ESD in wire segmentation and 99.05% F1 score in via detection on SEM images collected in real industrial applications, which outperforms the reference method[3] by 85% on ESD and 8% on F1 score, respectively.

### 5. REFERENCES

- Ronald Wilson, Navid Asadizanjani, Domenic Forte, and Damon L. Woodard, "Histogram-based auto segmentation: A novel approach to segmenting integrated circuit structures from sem images," 2020.
- [2] Bruno Machado Trindade, Eranga Ukwatta, Mike Spence, and Chris Pawlowicz, "Segmentation of integrated circuit layouts from scan electron microscopy images," in 2018 IEEE Canadian Conference on Electrical Computer Engineering (CCECE), 2018, pp. 1–4.
- [3] Tong Lin, Yiqiong Shi, Na Shu, Deruo Cheng, Xuenong Hong, Jingsi Song, and Bah Hwee Gwee, "Deep learning-based image analysis framework for hardware assurance of digital integrated circuits," in 2020 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA), 2020, pp. 1–6.
- [4] Xuenong Hong, Deruo Cheng, Yiqiong Shi, Tong Lin, and Bah Hwee Gwee, "Deep learning for automatic ic image analysis," in 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), 2018, pp. 1–5.
- [5] Ronald Wilson, Hangwei Lu, Mengdi Zhu, Domenic Forte, and Damon L. Woodard, "Refics: Assimilating data-driven paradigms into reverse engineering and hardware assurance on integrated circuits," *IEEE Access*, vol. 9, pp. 131955–131976, 2021.
- [6] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao, "Deep high-resolution representation learning for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3349– 3364, 2021.
- [7] Nahian Siddique, Sidike Paheding, Colin P. Elkin, and Vijay Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021.
- [8] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.
- [9] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, "Mask r-cnn," in 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988.

- [10] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [11] Simon Jégou, Michal Drozdzal, David Vazquez, Adriana Romero, and Yoshua Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 1175–1183.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
- [13] Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz, "Beyond accuracy, f-score and roc: A family of discriminant measures for performance evaluation," Berlin, Heidelberg, 2006, AI'06, p. 1015–1021, Springer-Verlag.