Imbalance Aware Lithography Hotspot Detection: A Deep Learning Approach

Haoyu Yang[1], Luyang Luo[1], Jing Su[2], Chenxi Lin[2], **Bei Yu**[1]

[1]The Chinese University of Hong Kong
[2]ASML Brion Inc.

Mar. 1, 2017

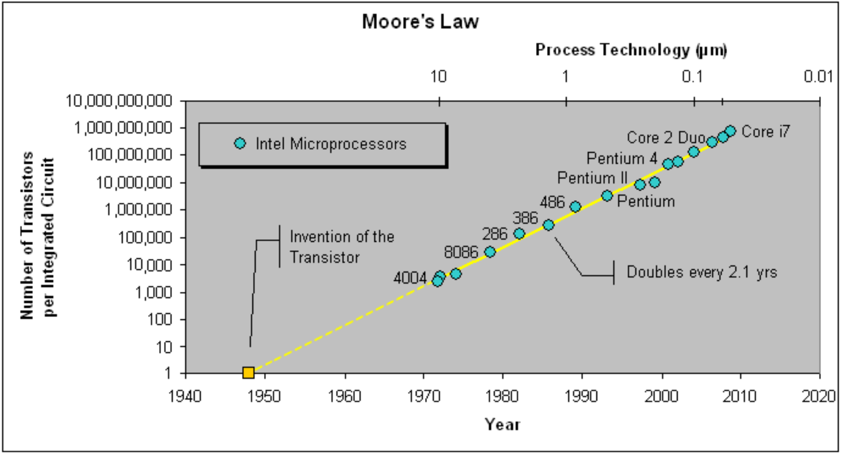
Outline

Outline

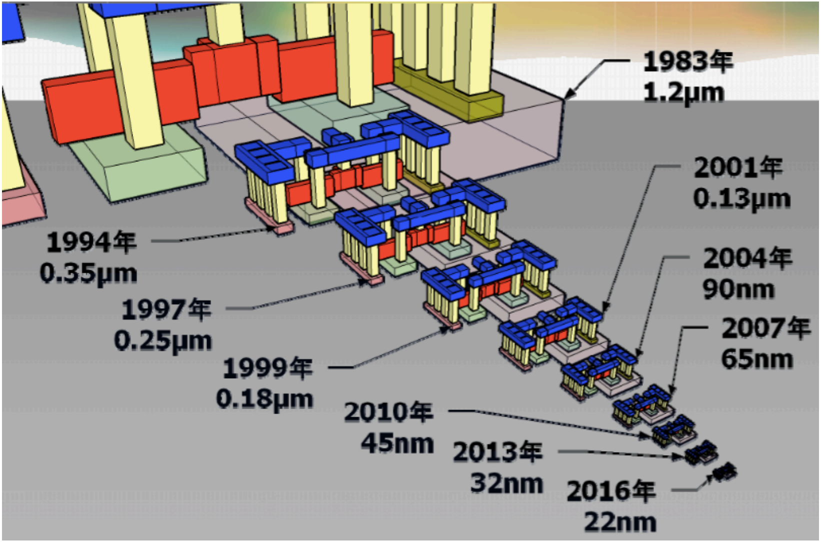
SPIE.

CONNECTING MINDS. ADVANCING LIGHT.

Moore's Law to Extreme Scaling

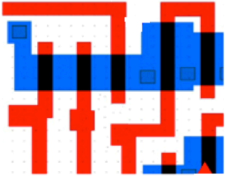


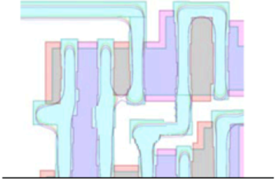
SPIE. CONNECTING MINDS. ADVANCING LIGHT.

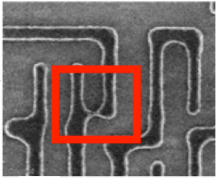


SPIE.
CONNECTING MINDS.
ADVANCING LIGHT.

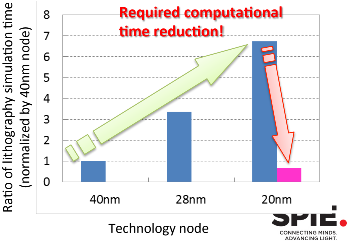
Lithography Hotspot Detection



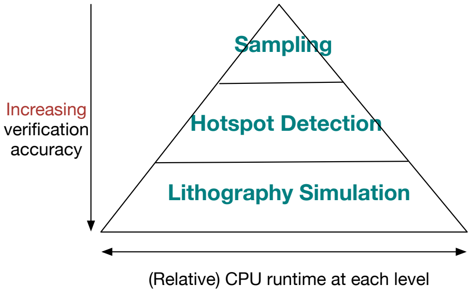


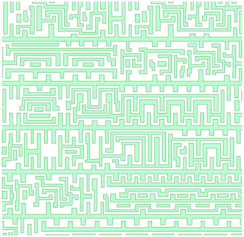


- What you see $\neq$ what you get
- Even w. RET: OPC, SRAF, MPL
- Still hotspot: low fidelity patterns
- Simulations: extremely CPU intensive



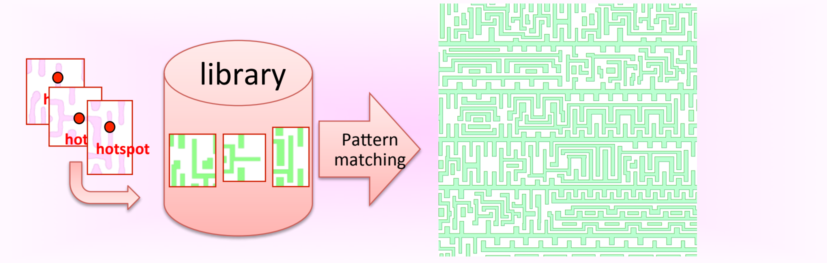
Layout Verification Hierarchy



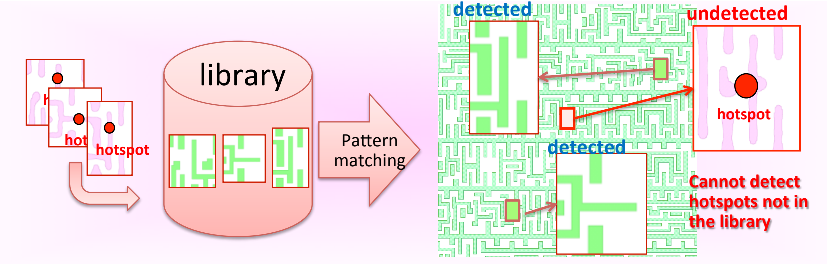


- **Sampling**:
  scan and rule check each region
- **Hotspot Detection**:
  verify the sampled regions and report potential hotspots
- **Lithography Simulation**:
  final verification on the reported hotspots

Pattern Matching based Hotspot Detection

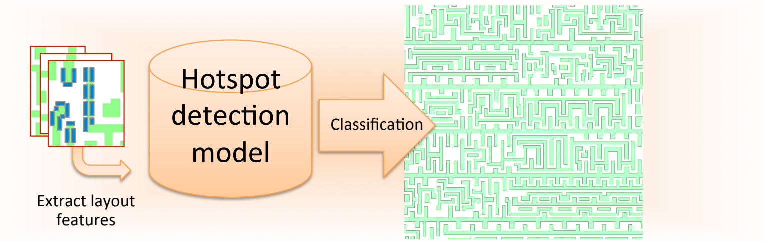


Pattern Matching based Hotspot Detection

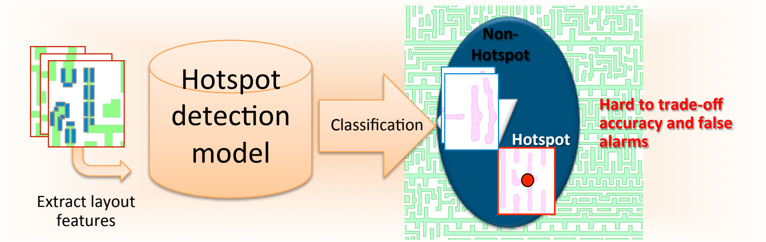


- Fast and accurate
- [Yu+,ICCAD'14] [Nosato+,JM3'14] [Su+,TCAD'15]
- Fuzzy pattern matching [Wen+,TCAD'14]
- Hard to detect non-seen pattern

Machine Learning based Hotspot Detection



Machine Learning based Hotspot Detection



- Predict new patterns
- Decision-tree, ANN, SVM, Boosting ...
- [Drmanac+,DAC'09] [Ding+,TCAD'12] [Yu+,JM3'15] [Matsunawa+,SPIE'15] [Yu+,TCAD'15][Zhang+,ICCAD'16]
- Crafted features are not satisfactory
- Hard to handle ultra-large datasets.

SPIE. CONNECTING MINDS. ADVANCING LIGHT.

Why Deep Learning?

- **Feature Crafting v.s. Feature Learning**
  Although prior knowledge is considered during manually feature design, information loss is inevitable.
  Feature learned from mass dataset is more reliable.
- **Scalability**
  With shrinking down circuit feature size, mask layout becomes more complicated. Deep learning has the potential to handle ultra-large-scale instances while traditional machine learning may suffer from performance degradation.
- **Mature Libraries**
  `Caffe` [Jia+,ACMMM'14] and `Tensorflow` [Martin+,TR'15]

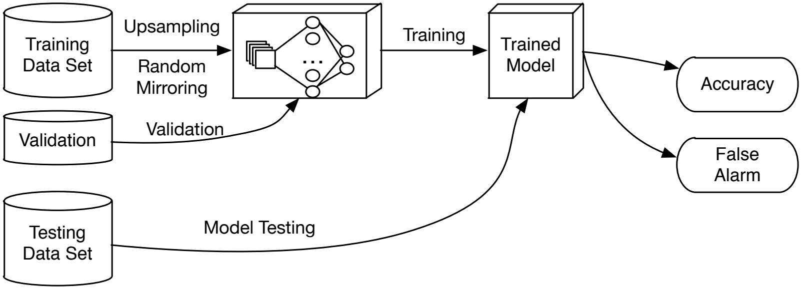
Hotspot-Oriented Deep Learning

Deep Learning has been widely appied in object recognition tasks.
Nature of mask layout impedes the availability of existing frameworks.

- **Imbalanced Dataset**
  Lithographic hotspots are always the minority.
- **Larger Image Size**
  Effective clip region ($> 1000 \times 1000$ pixels) is much larger than the image size in traditional computer vision problems.
- **Sensitive to Scaling**
  Scaling of mask layout patterns modifies its attributes.

SPIE. CONNECTING MINDS. ADVANCING LIGHT.

Deep Learning based Hostpot Detection Flow

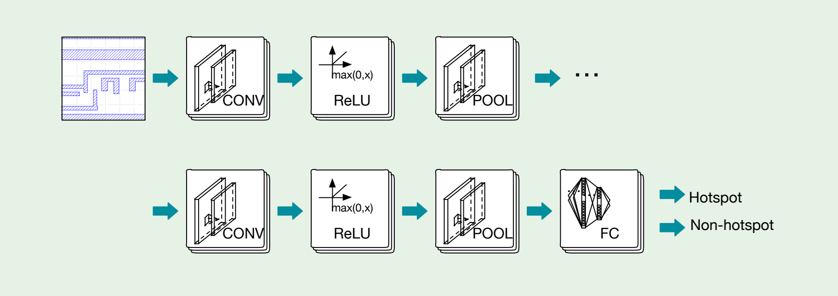


Outline

SPIE.
CONNECTING MINDS. ADVANCING LIGHT.

CNN Architecture Overview

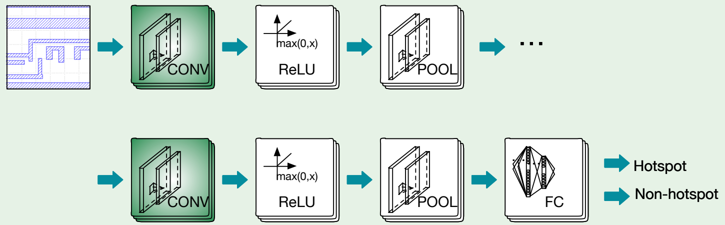
- Convolution Layer
- Rectified Linear Unit (ReLU)
- Pooling Layer
- Fully Connected Layer



Convolution Layer

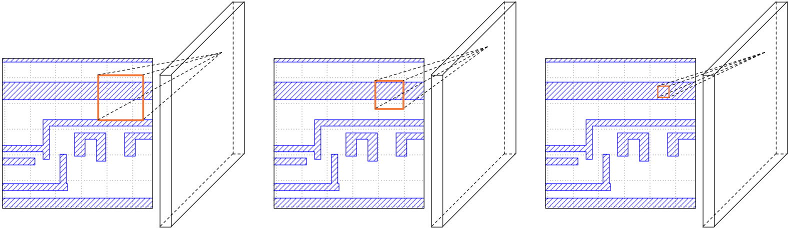
Convolution Operation:

$$\mathbf{I} \otimes \mathbf{K}(x, y) = \sum_{i=1}^{c} \sum_{j=1}^{m} \sum_{k=1}^{m} \mathbf{I}(i, x - j, y - k)\mathbf{K}(j, k)$$



Convolution Layer (cont.)

Effect of different convolution kernel sizes:

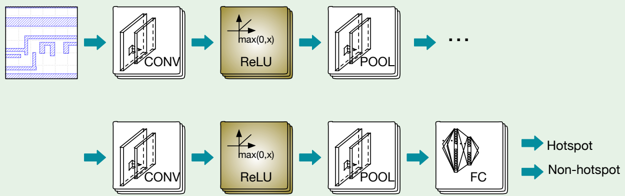


(a) 7×7 (b) 5×5 (c) 3×3

Kernel Size	Padding	Test Accuracy∗
7×7	3	87.50%
5×5	2	93.75%
3×3	1	**96.25%**

∗Stop after 5000 iterations.

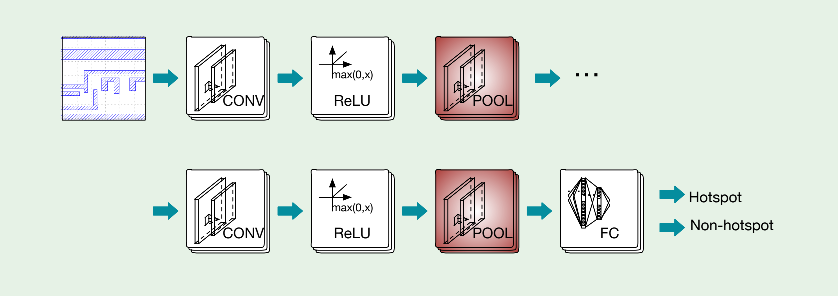
Rectified Linear Unit



- Alleviate overfitting with sparse feature map
- Avoid gradient vanishing problem

Activation Function	Expression	Validation Loss
ReLU	$\max\{x, 0\}$	**0.16**
Sigmoid	$\frac{1}{1+\exp(-x)}$	87.0
TanH	$\frac{\exp(2x)-1}{\exp(2x)+1}$	0.32
BNLL	$\log(1+\exp(x))$	87.0
WOAF	NULL	87.0

Pooling Layer



- Extracts the local region statistical attributes in the feature map

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

MAXPOOL

6	8
14	16

(a) max pooling

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

AVEPOOL

3.5	5.5
11.5	13.5

(b) avg pooling

Pooling Layer (cont.)

- Translation invarient (✘)
- Dimension reduction

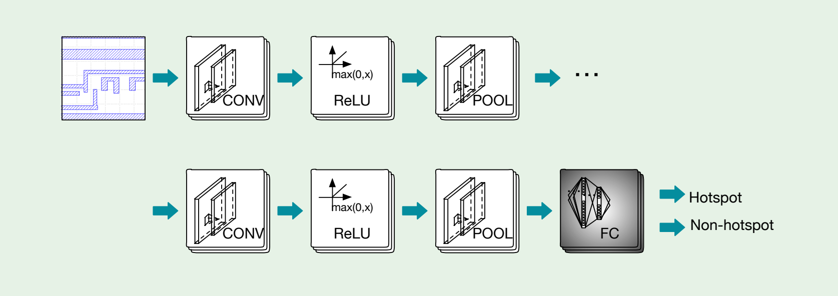
Effect of pooling methods:

Pooling Method	Kernel	Test Accuracy
Max	2×2	**96.25%**
Ave	2×2	**96.25%**
Stochastic	2×2	90.00%

SPIE. CONNECTING MINDS. ADVANCING LIGHT.

Fully Connected Layer

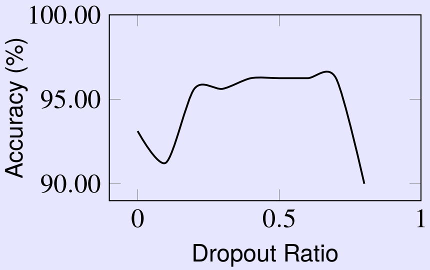
- Fully connected layer transforms high dimension feature maps into flattened vector.

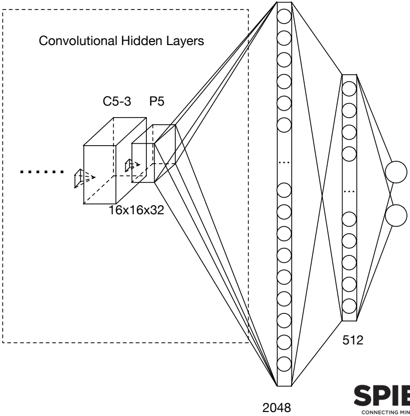


Fully Connected Layer (cont.)

- A percentage of nodes are dropped out (i.e. set to zero)
- avoid overfitting

Effect of dropout ratio:

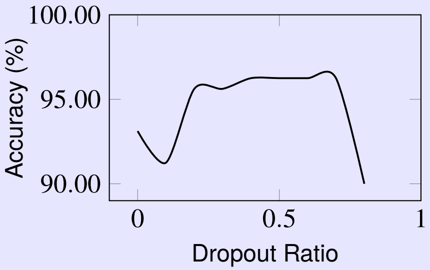


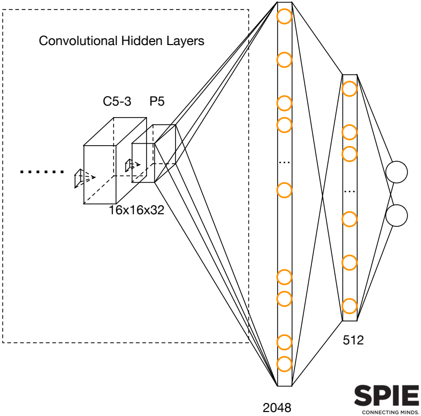


Fully Connected Layer (cont.)

- A percentage of nodes are dropped out (i.e. set to zero)
- avoid overfitting

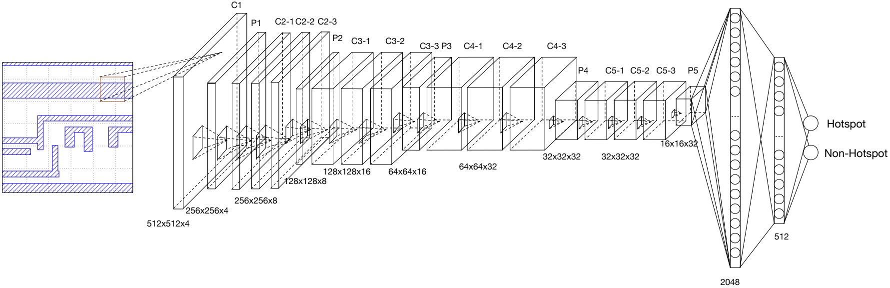
Effect of dropout ratio:





Architecture Summary

- Total 21 layers with 13 convolution layers and 5 pooling layers.
- A ReLU is applied after each convolution layer.



SPIE.
CONNECTING MINDS. ADVANCING LIGHT.

Architecture Summary

Layer	Kernel Size	Stride	Padding	Output Vertexes
Conv1-1	$2 \times 2 \times 4$	2	0	$512 \times 512 \times 4$
Pool1	2×2	2	0	$256 \times 256 \times 4$
Conv2-1	$3 \times 3 \times 8$	1	1	$256 \times 256 \times 8$
Conv2-2	$3 \times 3 \times 8$	1	1	$256 \times 256 \times 8$
Conv2-3	$3 \times 3 \times 8$	1	1	$256 \times 256 \times 8$
Pool2	2×2	2	0	$128 \times 128 \times 8$
Conv3-1	$3 \times 3 \times 16$	1	1	$128 \times 128 \times 16$
Conv3-2	$3 \times 3 \times 16$	1	1	$128 \times 128 \times 16$
Conv3-3	$3 \times 3 \times 16$	1	1	$128 \times 128 \times 16$
Pool3	2×2	2	0	$64 \times 64 \times 16$
Conv4-1	$3 \times 3 \times 32$	1	1	$64 \times 64 \times 32$
Conv4-2	$3 \times 3 \times 32$	1	1	$64 \times 64 \times 32$
Conv4-3	$3 \times 3 \times 32$	1	1	$64 \times 64 \times 32$
Pool4	2×2	2	0	$32 \times 32 \times 32$
Conv5-1	$3 \times 3 \times 32$	1	1	$32 \times 32 \times 32$
Conv5-2	$3 \times 3 \times 32$	1	1	$32 \times 32 \times 32$
Conv5-3	$3 \times 3 \times 32$	1	1	$32 \times 32 \times 32$
Pool5	2×2	2	0	$16 \times 16 \times 32$
FC1	–	–	–	2048
FC2	–	–	–	512
FC3	–	–	–	2

Outline

Introduction

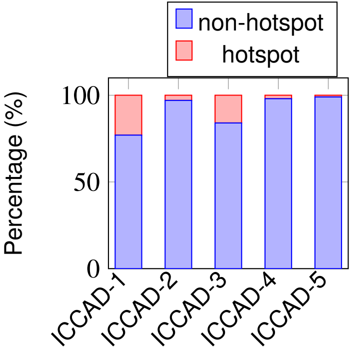
Network Architecture

Imbalance Aware Learning

Experimental Results

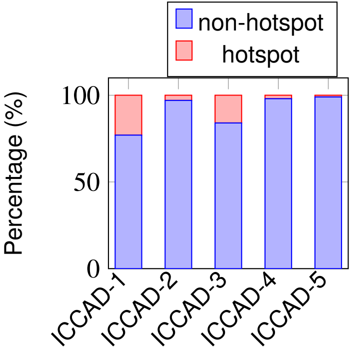
Minority Upsampling

Layout datasets are highly imbalanced as after resolution enhancement techniques (RETs) the lithographic hotspots are always the minority.



Minority Upsampling

Layout datasets are highly imbalanced as after resolution enhancement techniques (RETs) the lithographic hotspots are always the minority.

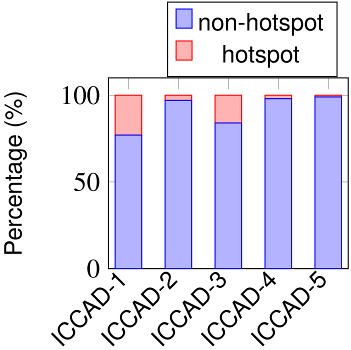


- Multi-label learning [Zhang+,IJCAI'15]
- Majority downsampling [Ng+,TCYB'15]
- Pseudo instance generation [He+,IJCNN'08]
  Artifically generated instances might not be available because of mask layout nature.

SPIE. CONNECTING MINDS. ADVANCING LIGHT.

Minority Upsampling

Layout datasets are highly imbalanced as after resolution enhancement techniques (RETs) the lithographic hotspots are always the minority.

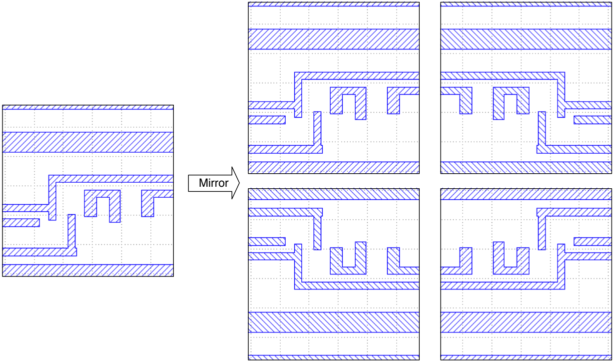


- Multi-label learning [Zhang+,IJCAI'15]
- Majority downsampling [Ng+,TCYB'15]
- Pseudo instance generation [He+,IJCNN'08]
  Artifically generated instances might not be available because of mask layout nature.
- Naïve upsampling (✓)
  1. Gradient descent
  2. Insufficient training samples

SPIE. CONNECTING MINDS. ADVANCING LIGHT.

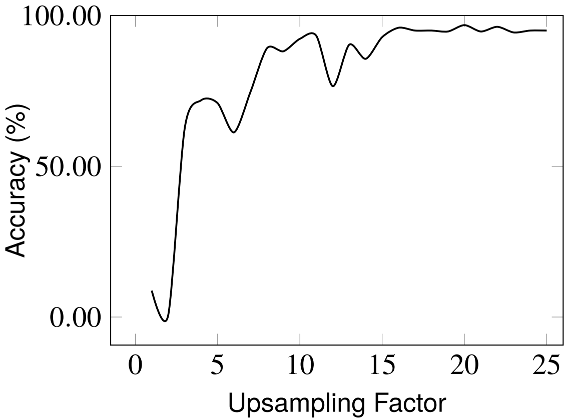
Random Mirror Flipping

- Before fed into neural network
- Each instance is taking one of 4 orientations
- Resolve insufficient data



SPIE. CONNECTING MINDS. ADVANCING LIGHT.

Effectiveness of Upsampling



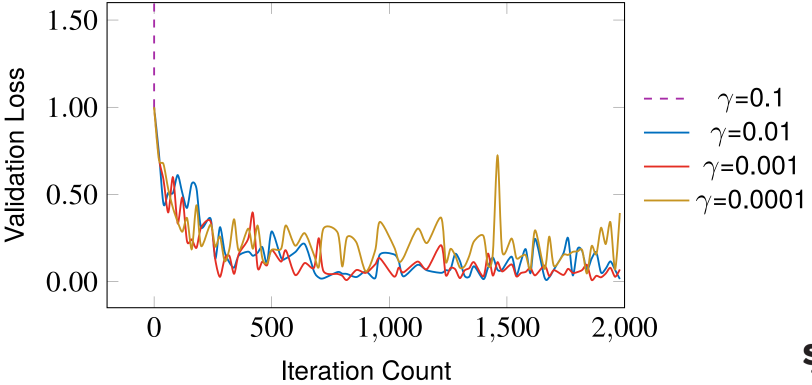
Validation performance does not show further improvement when the upsampling factor increases beyond a certain value.

SPIE.
CONNECTING MINDS. ADVANCING LIGHT.

Learning Rate

γ: defines how fast the neuron weights are updated

$$w_i = w_i - \gamma \frac{\partial l}{\partial w_i}.$$



Momentum and Weight Decay

- Momentum

  Physical meaning is involved into gradient descent.

$$v = \mu v - \gamma \frac{\partial l}{\partial w_i},$$
$$w_i = w_i + v.$$

- Weight Decay

  An alternative to achieve L_2 regularization on neuron weights.

$$v = \mu v - \gamma \frac{\partial l}{\partial w_i} - \gamma \lambda w_i,$$
$$w_i = w_i + v.$$

SPIE.
CONNECTING MINDS. ADVANCING LIGHT.

Momentum and Weight Decay (cont.)

- Momentum Effects:

μ	Learning Rate	Validation Loss
0.5	0.001	0.21
0.9	0.001	0.22
0.95	0.001	0.21
0.99	0.001	**0.16**

- Weight Decay Effects:

λ	Learning Rate	Momentum	Validation Loss
10^{-3}	0.001	0.99	0.95
10^{-4}	0.001	0.99	1.19
10^{-5}	0.001	0.99	0.37
10^{-6}	0.001	0.99	**0.2**

SPIE.
CONNECTING MINDS. ADVANCING LIGHT.

Weight Initialization

The weight initialization procedure determines what initial values assigned to each neuron before the gradient descent update starts.

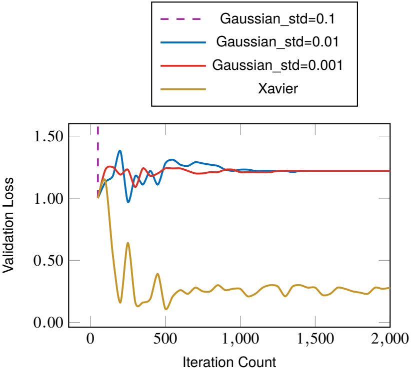
- Random Gaussian (✘)
  Cannot guarantee input &
  output have similar variance.

Weight Initialization

The weight initialization procedure determines what initial values assigned to each neuron before the gradient descent update starts.

- Random Gaussian (✗)
  Cannot guarantee input & output have similar variance.
- Xavier [Xavier+,AISTATS'10]
  Initialized weights are determined by input node number.

$$\hat{V}(w_i) = \frac{1}{N}.$$



Outline

SPIE.
CONNECTING MINDS. ADVANCING LIGHT.

Experimental Setup

- Based on `Caffe` [Jia+,ACMMM'14]
- Evaluated on ICCAD-2012 CAD contest benchmark

Evaluation metrics:

Accuracy

The ratio between the number of correctly detected hotspot clips and the number of all hotspot clips.

ODST

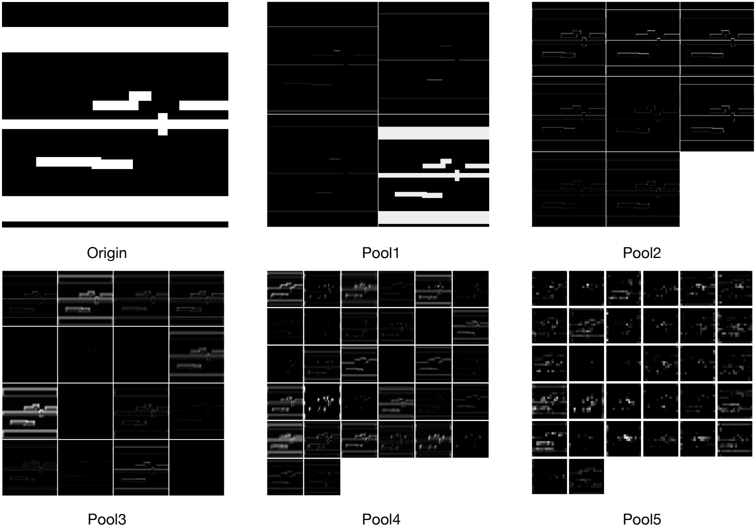
The sum of all lithographic simulation time for false alarm† and the deep learning model testing time.

$$\text{ODST} = \text{Test Time} + 10\text{s} \times \# \text{ of False Alarm}$$

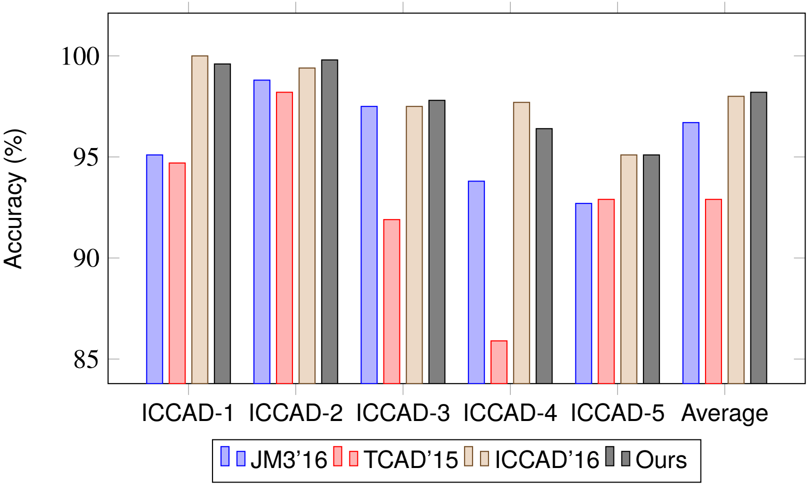
†False alarm: the number of non-hotspot clips that are reported as hotspots by detector.

SPIE. CONNECTING MINDS. ADVANCING LIGHT.

Layer Visualization



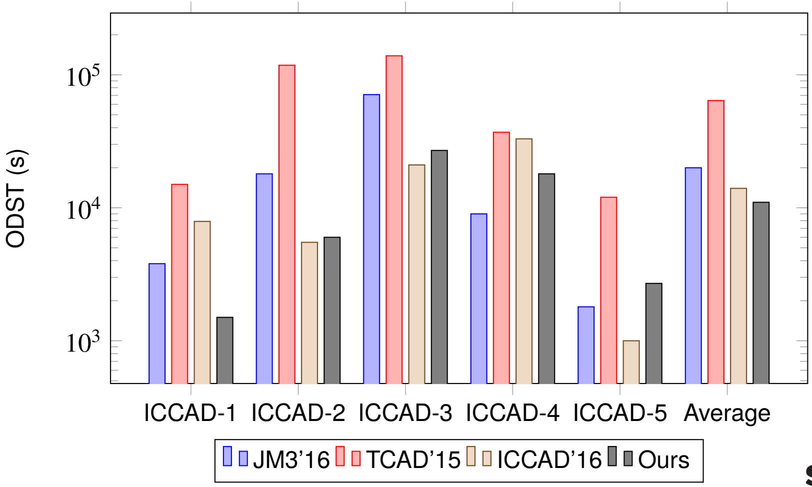
Compare Accuracy with State-of-the-Art‡



‡JM3'16: CNN based; TCAD'15: SVM based; ICCAD'16: Boosting based.

Compare ODST with State-of-the-Art

- Improve the performance of ODST by at least 24.80% on average.

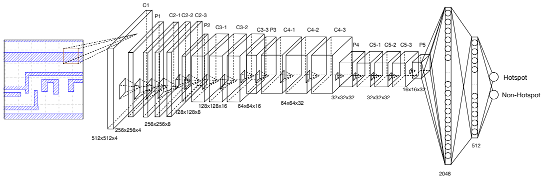


JM3'16: CNN based; TCAD'15: SVM based; ICCAD'16: Boosting based.

Conclusion

We explore the feasibility of deep learning as an alternative approach for hotspot detection.

- Hotspot-detection-oriented hyper-parameter tuning
- Imbalance Issue: Upsampling & Random mirror flipping
- Outperform state-of-the-art solutions

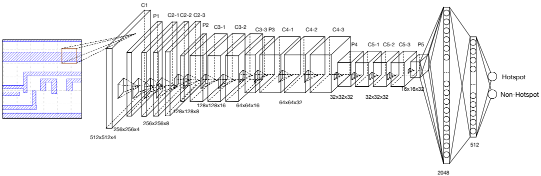


SPIE. CONNECTING MINDS. ADVANCING LIGHT.

Conclusion

We explore the feasibility of deep learning as an alternative approach for hotspot detection.

- Hotspot-detection-oriented hyper-parameter tuning
- Imbalance Issue: Upsampling & Random mirror flipping
- Outperform state-of-the-art solutions



Future Works

- Test on larger scale test cases
- Further simplify architecture to speedup
- Seek other VLSI layout applications (e.g., OPC, SRAF)

Thank You

Haoyu Yang (hyyang@cse.cuhk.edu.hk)

Luyang Luo (lyluo4@cse.cuhk.edu.hk)

Jing Su (jing.su@asml.com)

Chenxi Lin (chenxi.lin@asml.com)

Bei Yu (byu@cse.cuhk.edu.hk)