TEXAS
The University of Texas at Austin

High-Definition Routing Congestion Prediction for Large-Scale FPGAs

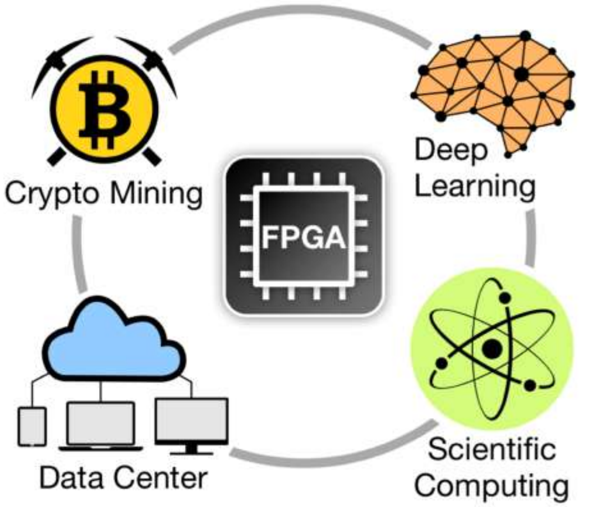
Mohamed Baker Alawieh[1], Wuxi Li[1], **Yibo Lin**[2], Love Singhal[3], Mahesh Iyer[3] and David Z. Pan[1]

[1]ECE Department, University of Texas at Austin

[2]CS Department, Peking University

[3]Intel Corporation, USA

FPGA Routing Congestion Prediction



Field Programmabe Gate Arrays

High Energy Efficiency
Good Reprogrammability
Rapidly Growing Capacity

FPGA Placement

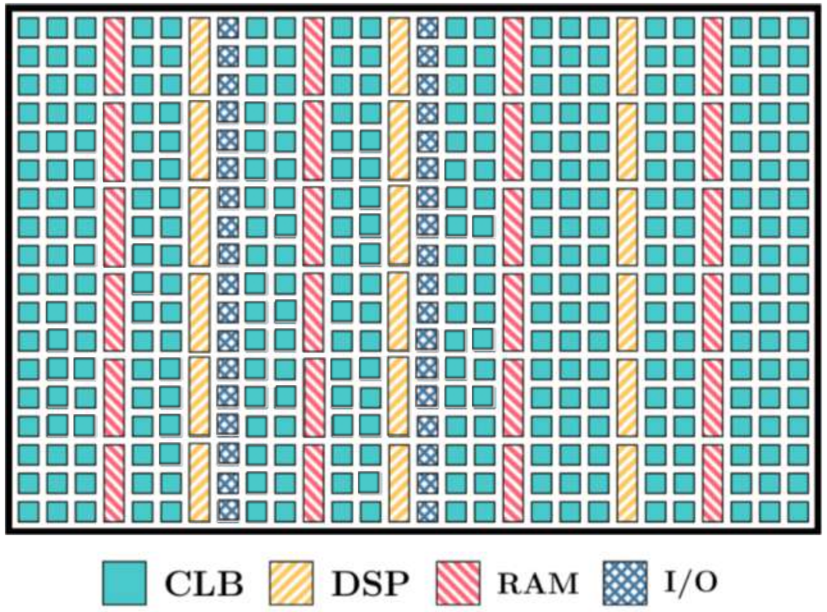
Has a significant impact on FPGA routing quality

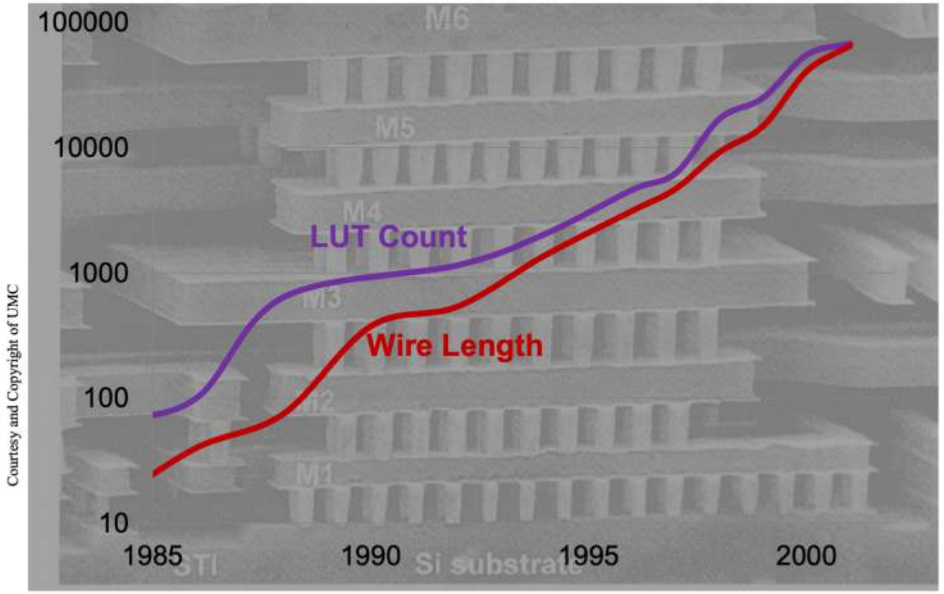
Routability Aware

Incorporates congestion prediction into the placement process

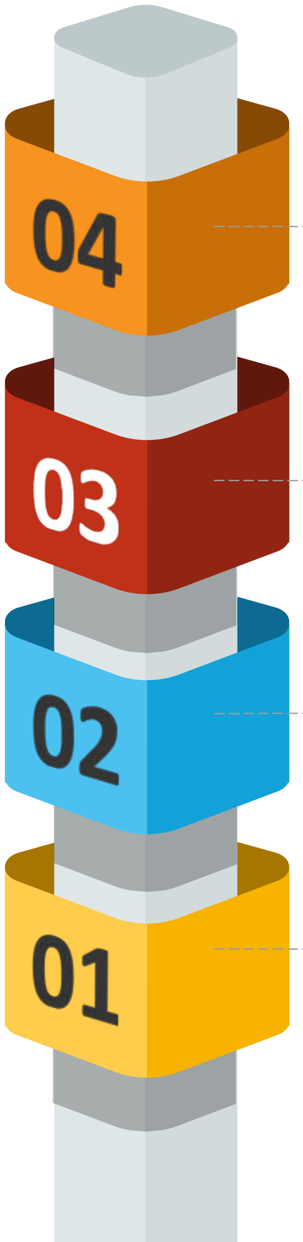
Congestion

Primitive congestion prediction techniques have demonstrated significant impact on routing quality





Conventional Approaches



GAN-Based

Predicts congestion based on placement
Cannot handle industrial-size designs
[Yu+, DAC'19]

RouteNet

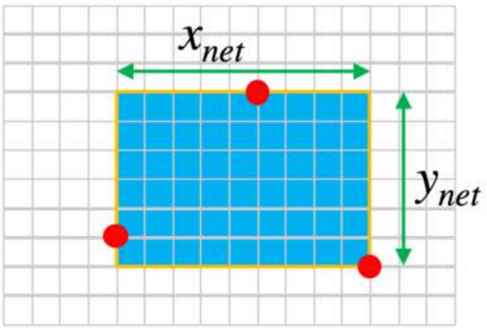
Predicts congestion hotspot
Design rule violation detection
[Xie+, ICCAD'18]

Regression-based Prediction

Congestion prediction based on global routing info
[Pui+, ICCAD'17]

RUDY

Bounding box-based routing estimation
Overestimates the routing demand
[Spindler+, DATE'07]



Conditional GANs for Image Translations [Isola+, CVPR 2017]

GANS

Generative Adversarial Networks
Generate Images from a distribution

CGANS

Conditional GANs
Generate an image based on input

Image Translation

CGANs can be used for the task
Apply domain transfer

Take image from one domain and generate output in another

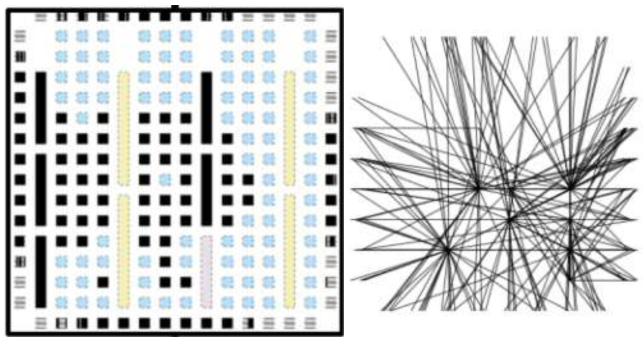
During training, pairs of matched images are used

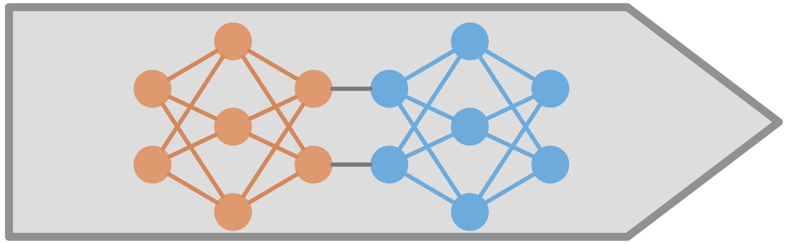
[cartoon credit: Gall, 18, dzone.com]

GAN-based Congestion Estimation

[Yu+, DAC'19]

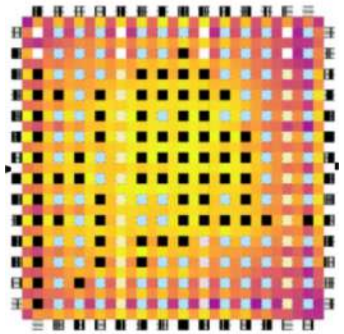
Placement and Netlist Information





CGAN-Based Image Translation

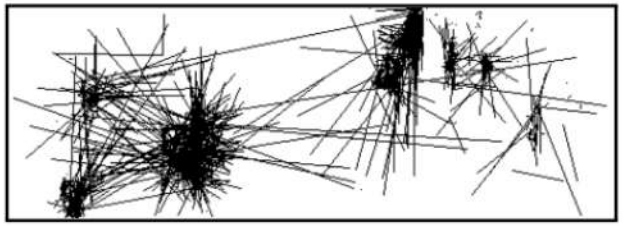
Congestion Map



Features

Uses VTR academic tool
Works for small designs only

Netlist information is encoded using flying lines
*For a large design with over 700K nets



Only 5K nets out of 700K shown

All 700K nets shown

This representation becomes obsolete for large designs

GAN Model

pix2pix model [Isola+, CVPR 2017]
Limited resolution 256x256
Cannot handle large-scale FPGAs

High-Definition Routing Prediction for Large FPGAs

GAN Model

pix2pix model [Isola+, CVPR 2017]
Limited resolution 256x256
Cannot handle large-scale FPGAs

Virtex UltraScale+ VU19 has ~663K CLB slices

Use a high definition image translation model
Handle resolution up to 4000x1000

Features

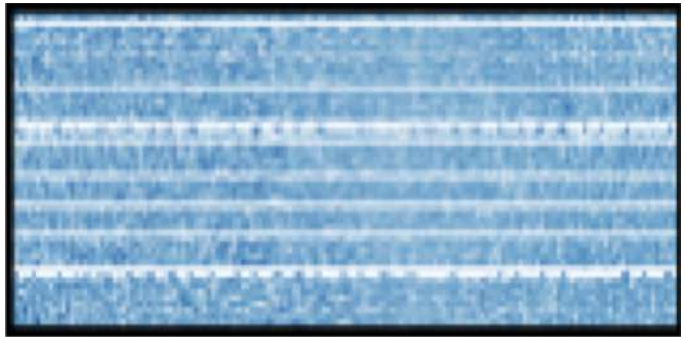
Uses VTR academic tool
Works for small designs only

Novel feature encoding for placement and netlist
Use different channels of input image

Input Features Encoding

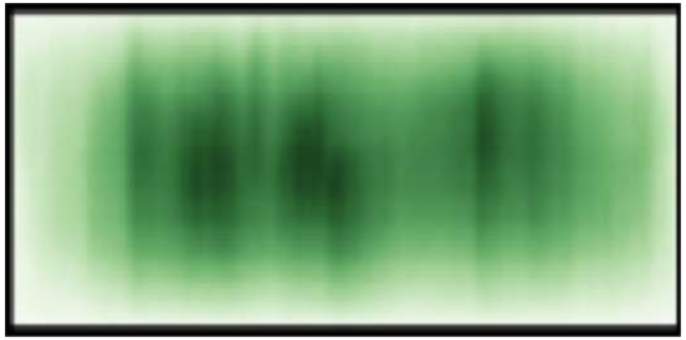
Pin Density

Reflects placement information
Encoded on the blue channel



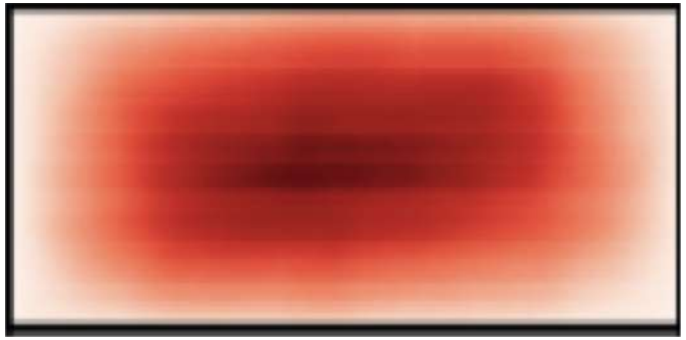
Vertical Demand

Estimtes vertical routing demand
Computed analogous to RUDY
Encoded on green channel

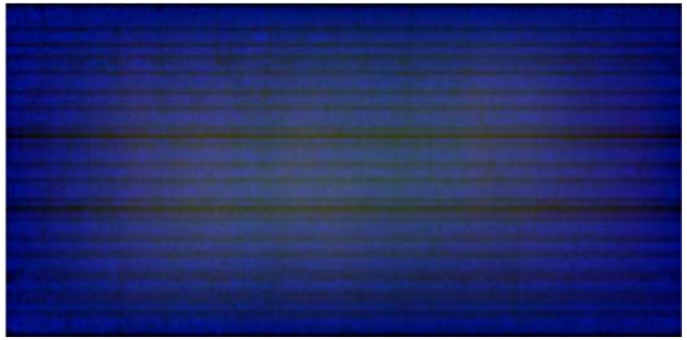


Horizontal Demand

Estimtes vertical routing demand
Computed analogous to RUDY
Encoded on red channel



Resulting RGB image



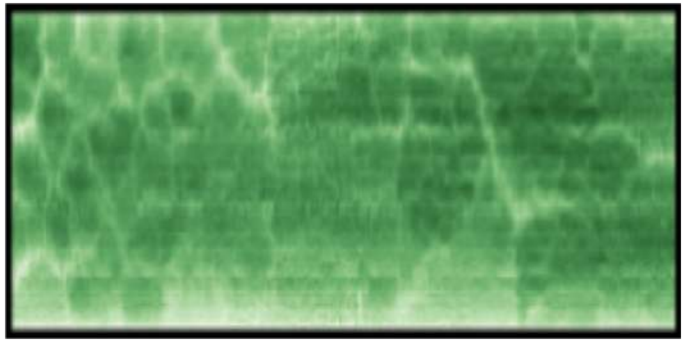
Output Features Encoding

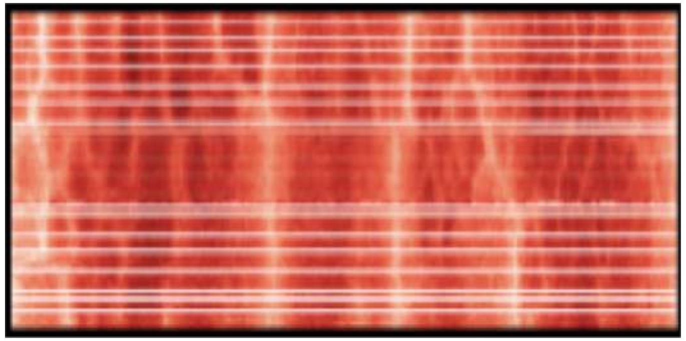
Vertical Routing

Routing congestion along the vertical direction

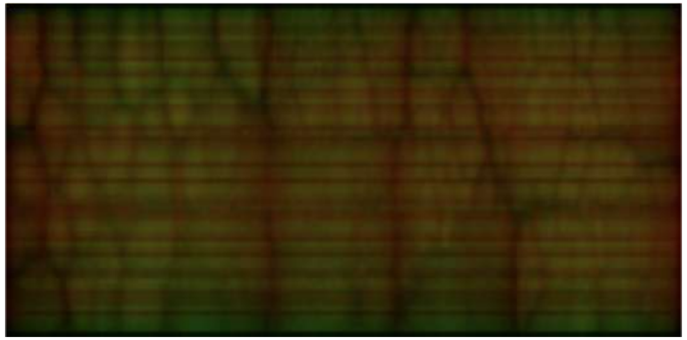
Horizontal Routing

Routing congestion along the horizontal direction





Resulting RGB image



Blue channel left empty

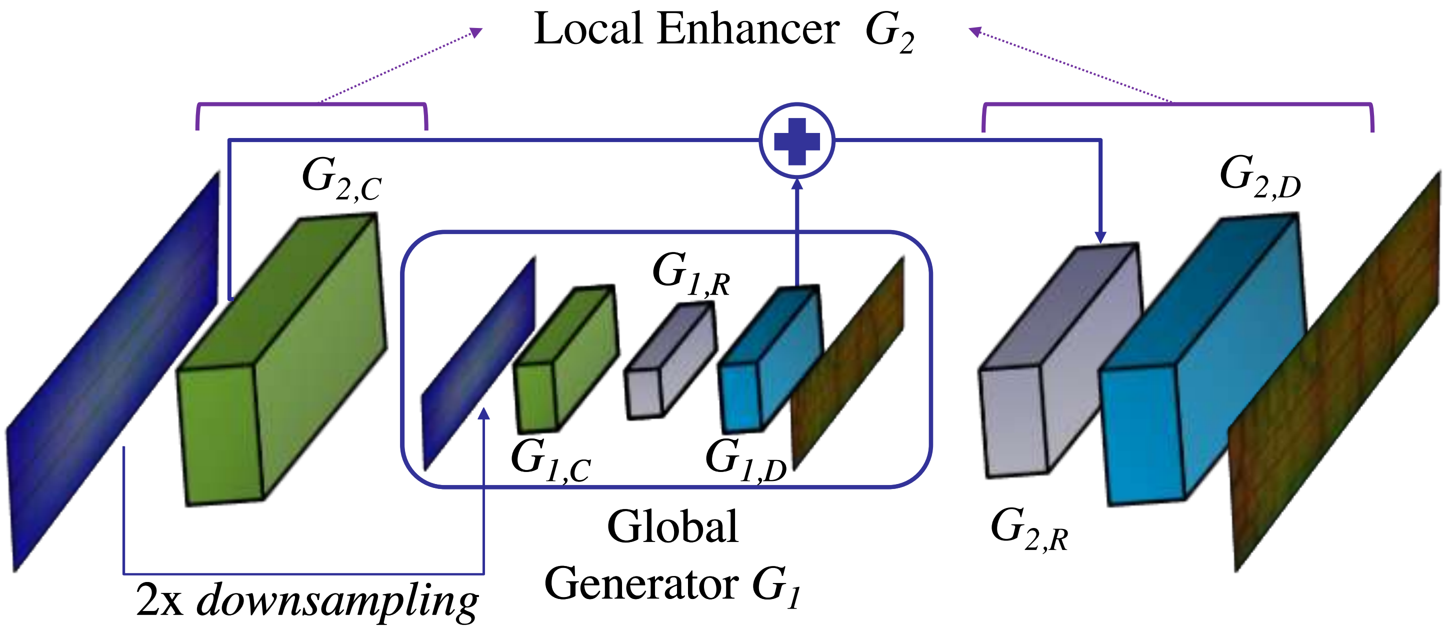
High Definition Image Translation

Generator Design

Dual generator architecture
For high resolution generation

Global Generator (G_1):
Performs the core translation
Works at half desired resolution

Local Enhancer (G_2):
Generates high resolution images
Fine-tunes details in the image



High Definition Image Translation

pix2pixHD [Wang+, CVPR'18]

Generator Design

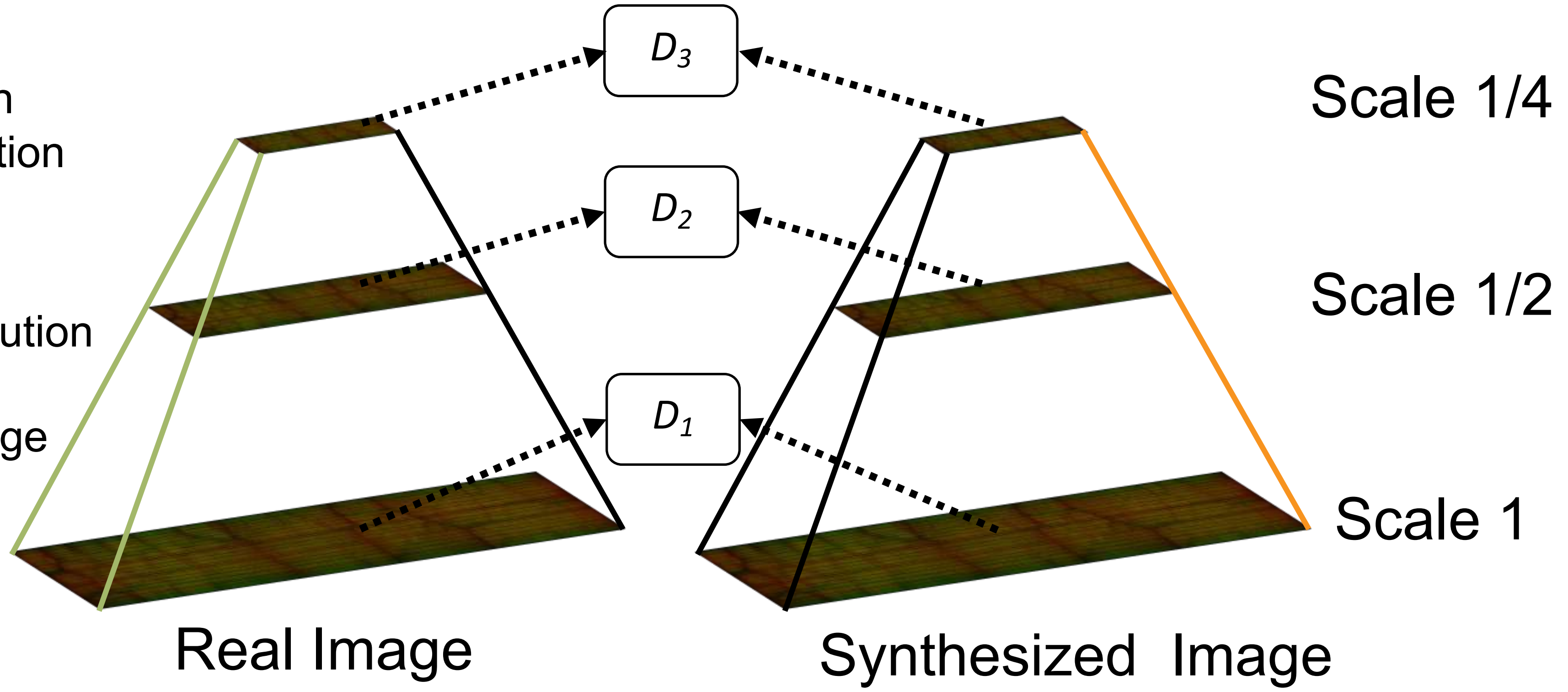
Dual generator architecture
For high resolution generation

Global Generator (G_1):
Performs the core translation
Works at half desired resolution

Local Enhancer (G_2):
Generates high resolution images
Fine-tunes details in the image

Discriminator Design

Three level discrimination



High Definition Image Translation

pix2pixHD [Wang+, CVPR'18]

Generator Design

Dual generator architecture
For high resolution generation

Discriminator Design

Three level discrimination

Loss Function

GAN Loss
Feature Mapping loss

Global Generator (G_1):
Performs the core translation
Works at half desired resolution

Local Enhancer (G_2):
Generates high resolution images
Fine-tunes details in the image

$$\min_{G} \max_{D_1, D_2, D_3} \sum_{k=1,2,3} \mathcal{L}_{GAN}(G, D_k) + \lambda \mathcal{L}_{FM}(G, D_k)$$

$$\mathcal{L}_{GAN}(G, D_k) = \mathbb{E}_{x,y}[\log D_k(x, y)] + \mathbb{E}_x[\log(1 - D_k(x, G(x)))]$$

$$\mathcal{L}_{FM}(G, D_k) = \mathbb{E}_{x,y} \sum_{i=1}^{T} ||D_k^{(i)}(x, y) - D_k^{(i)}(x, G(x))||_1$$

Experimental Setup

Benchmark

ISPD 2016
Placement: elfPlace [Li+, ICCAD'19]
Routing: NCTU-GR [Liu+, TCAD'13]

For each design:
200 placements are generated
Placements are routed
Congestion maps obtained

Training Setup

Train 12 different models
11 for train, 1 for test

Comparisons:
1.GAN-Based [Yu+, DAC19]
- Updated features
- Proper scaling

Compa
2.RUD

Benchmark	#LUT	#FF	#RAM	#DSP	#Ctrl Set
FPGA-1	50K	55K	0	0	12
FPGA-2	100K	66K	100	100	121
FPGA-3	250K	170K	600	500	1281
FPGA-4	250K	172K	600	500	1281
FPGA-5	250K	174K	600	500	1281
FPGA-6	350K	352K	1000	600	2541
FPGA-7	350K	355K	1000	600	2541
FPGA-8	500K	216K	600	500	1281
FPGA-9	500K	366K	1000	600	2541
FPGA-10	350K	600K	1000	600	2541
FPGA-11	480K	363K	1000	400	2091
FPGA-12	500K	602K	600	500	1281
Resources	538K	1075K	1728	768	N/A

Evaluation Metrics

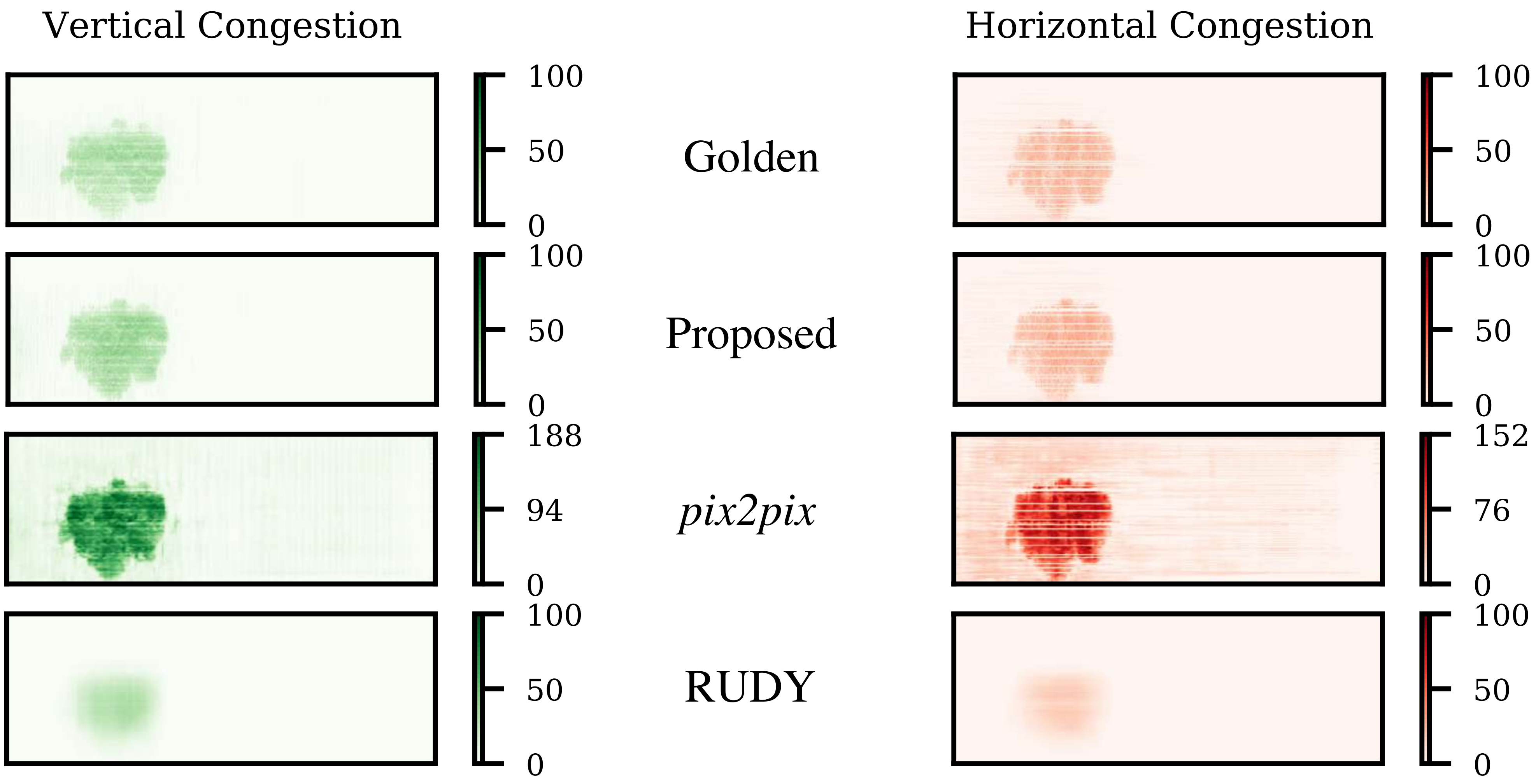
NRMS:
Normalized root mean square

SSIM:
Structural similarity index

EMD:
Earth moving distance
Difference in pixel distributions

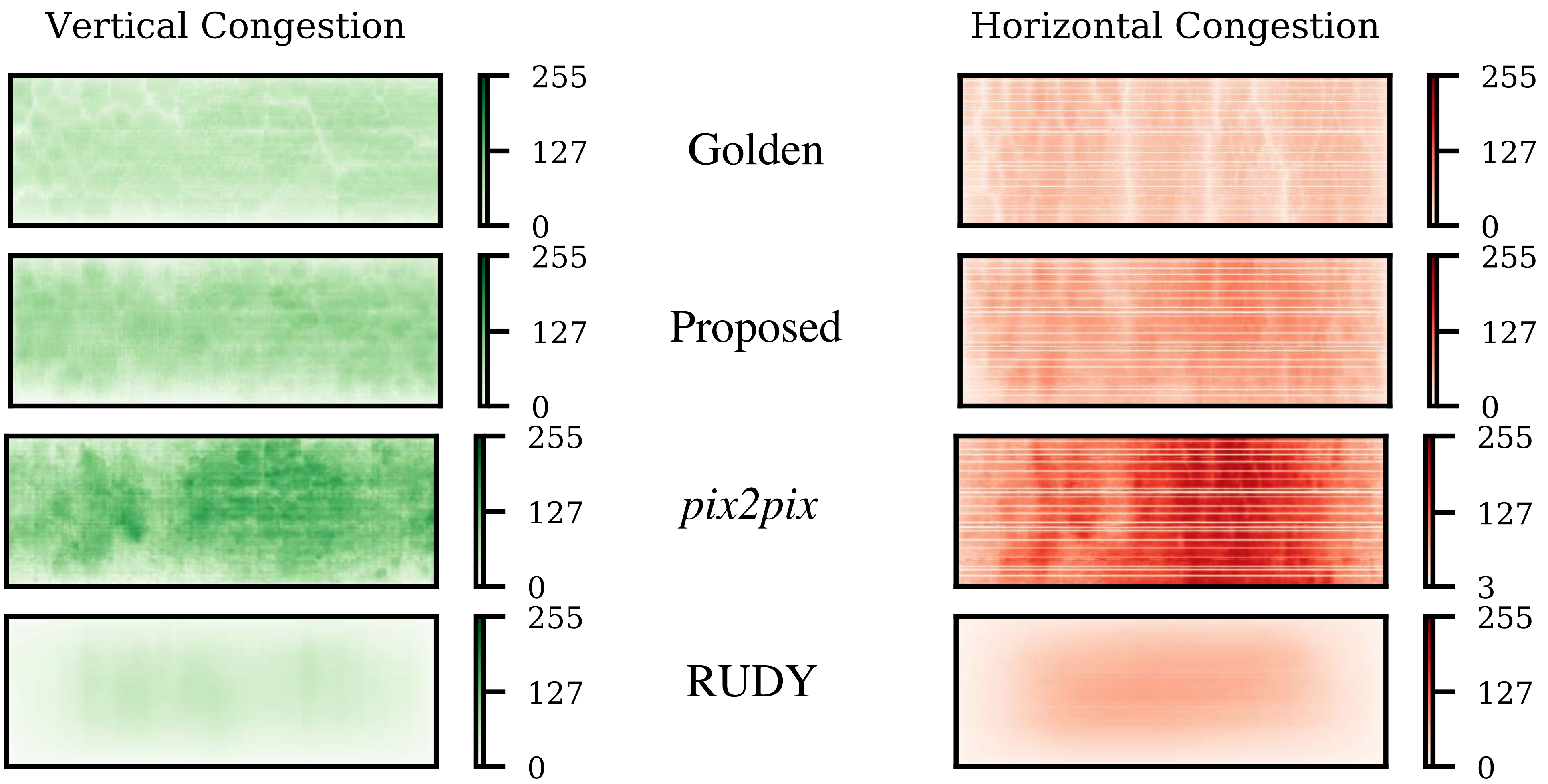
Sample Results – FPGA 02

RUDY ~ [Spindler+, DATE'07]
pix2pix ~ [Yu+, DAC'19]*



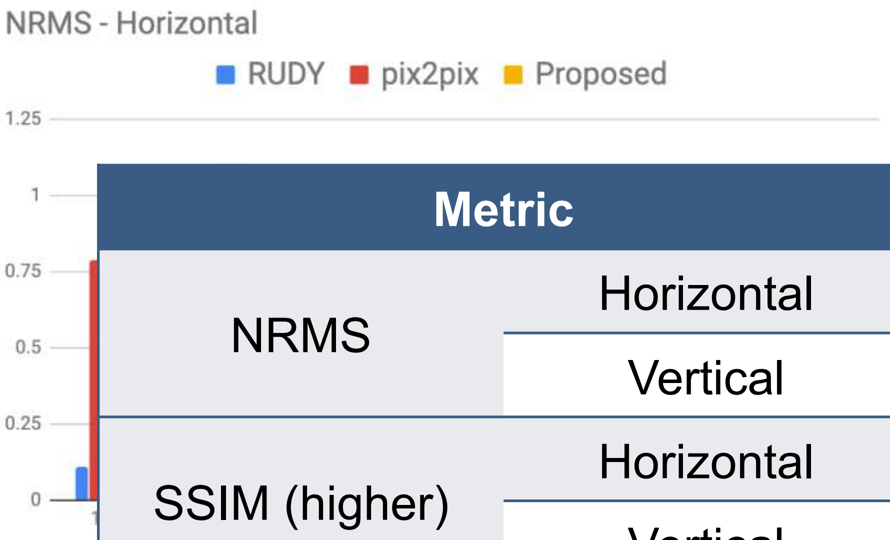
Sample Results – FPGA 08

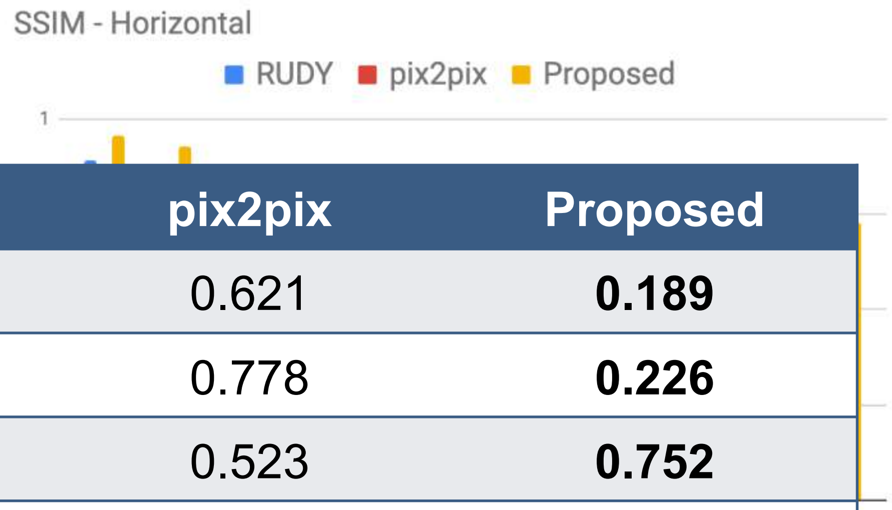
RUDY ~ [Spindler+, DATE'07]
pix2pix ~ [Yu+, DAC'19]*



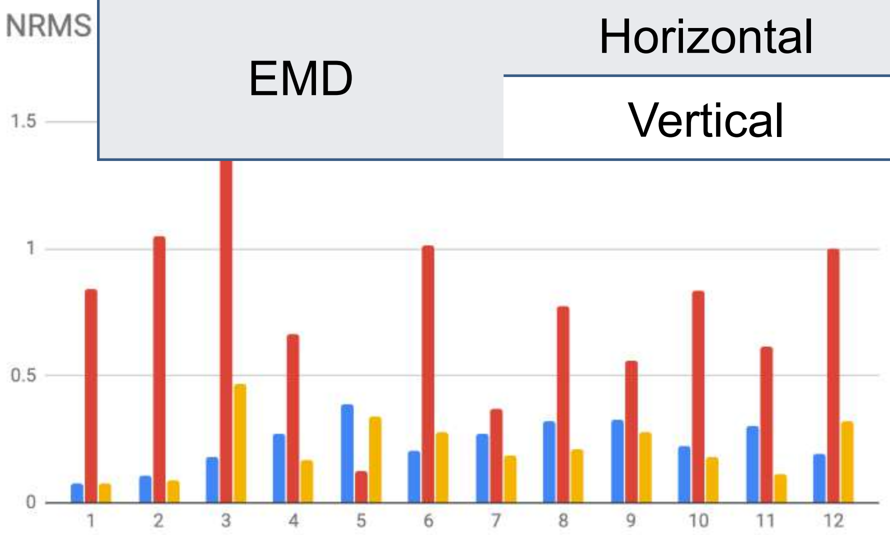
Quantitative Comparison

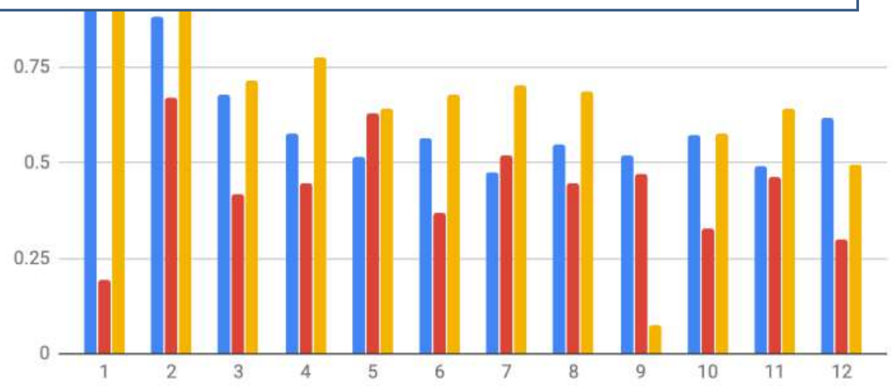
RUDY ~ [Spindler+, DATE'07]
pix2pix ~ [Yu+, DAC'19]*





Metric		RUDY	pix2pix	Proposed
NRMS	Horizontal	0.241	0.621	**0.189**
	Vertical	0.239	0.778	**0.226**
SSIM (higher)	Horizontal	0.407	0.523	**0.752**
	Vertical	0.616	0.439	**0.656**
EMD	Horizontal	0.162	0.225	**0.137**
	Vertical	0.137	0.233	**0.127**





Model Application

In Placement

Models were used for routability estimation within elfPlaceF replacing RUDY

Design	Full Routing Capacity		
	Rudy	Proposed	Imp
FPGA-1	336117	336117	0.00%
FPGA-2	691618	691618	0.00%
FPGA-3	3062734	3062734	0.00%
FPGA-4	5550659	5551473	-0.01%
FPGA-5	10538770	9797007	7.04%
FPGA-6	5773333	5773333	0.00%
FPGA-7	9182199	9163640	0.20%
FPGA-8	9053192	9053192	0.00%
FPGA-9	11641853	11635870	0.05%
FPGA-10	5515319	5515319	0.00%
FPGA-11	11777500	11757650	0.16%
FPGA-12	6235694	6235694	0.00%

FPGA-5 is the most congested design

Model Application

elfPlace [Li+, ICCAD'19]

In Placement

Models were used for routability estimation within elfPlaceF replacing RUDY

ROUTED WL REDUCTION

Design	Full Routing Capacity			Reduced Routing Capacity		
	Rudy	Proposed	Imp	Rudy	Proposed	Imp
FPGA-1	336117	336117	0.00%	336117	336117	0.00%
FPGA-2	691618	691618	0.00%	691618	691618	0.00%
FPGA-3	3062734	3062734	0.00%	3062734	3062734	0.00%
FPGA-4	5550659	5551473	-0.01%	5557608	5551473	0.11%
FPGA-5	10538770	9797007	7.04%	N/A	N/A	N/A
FPGA-6	5773333	5773333	0.00%	5777149	5773333	0.07%
FPGA-7	9182199	9163640	0.20%	9199730	9163640	0.39%
FPGA-8	9053192	9053192	0.00%	9055093	9055093	0.00%
FPGA-9	11641853	11635870	0.05%	11652436	11635870	0.14%
FPGA-10	5515319	5515319	0.00%	5515319	5515319	0.00%
FPGA-11	11777500	11757650	0.16%	11877778	11757650	1.01%
FPGA-12	6235694	6235694	0.00%	6224962	6235694	-0.17%

FPGA-5 is the most congested design

Conclusions

- We propose an accurate FPGA routing congestion estimation framework based on high-definition image translation

- Our proposed approach demonstrate superior accuracy compared to state-of-the-art techniques

- Our proposed approach results in up to 7% reduction in routed wirelength

Future Work

- Further improve feature representation
  - Preserve original connectivity information in feature encoding

- Develop new placement algorithm built around such accurate congestion estimation

- Extend the application to ASIC