Late Breaking Results: A Neural Network that Routes ICs

Dmitry Utyamishev
University of Illinois at Chicago
Chicago, IL
dutyam2@uic.edu

Inna Partin-Vaisband
University of Illinois at Chicago
Chicago, IL
vaisband@uic.edu

ABSTRACT

A global router is proposed that learns from routed circuits and autonomously routes unseen layouts. The uniqueness of this approach is in redefining the global routing as a classical image-to-image processing problem. The imaging problem is efficiently solved with a deep learning system, comprising a variational autoencoder and custom loss function. This fundamentally new routing method provides a natural way for global routing parallelization. The deep router is designed, trained, and tested on an unseen 64×64 ISPD'98 benchmark circuit. The test results yield 3.2% decrease in routability and over 5X speedup in runtime as compared with the state-of-the-art FastRoute router.

KEYWORDS

Global routing; deep learning; variational autoencoder; FastRoute.

1 Introduction

Global routing is a fundamental step of physical integrated circuit (IC) design which generates routing paths within a layout (i.e., grid of rectangular global routing tiles) by connecting pins of placed IC components with a net within specific technology limitations. Optimal global routing is a NP-complete problem. To mitigate the design complexity, approximation heuristics have been proposed, yielding sub-optimal routing in reasonable time. Traditional routers typically exhibit the following flow. First, the individual nets are decomposed into multiple pin-to-pin routing objectives and each objective is solved with a pathfinding algorithm. Next, rip-up and reroute based algorithms are exploited for optimizing the initial routing solution within certain congestion and design constraints. A primary limitation of existing global routers is the series nature of the underlying algorithms which makes parallelization on massively parallel multi-processor hardware less practical.

传统依赖串行形成block

Statistical and probabilistic approaches (i.e., reinforcement ML, deep ML, or genetic algorithms) have also been proposed to enhance the traditional global routing flow [1]-[4]. To the best of the authors knowledge, all the existing learning-enhanced solutions use the standard global routing design flow, with one or several ML-enhanced design stages. In some approaches, congestion maps are predicted and used for a more effective rip-and-reroute iterative wirelength optimization [1,2]. Replacing the whole optimization stage with a genetic algorithm has also been explored [1,3]. Another existing approach is to use reinforcement learning based pathfinding [1,4] for a more efficient initialization of the routing paths. Note, that similar to the traditional routers, all these approaches exhibit the fundamental design limitations: (i) iterative nature of the routing process, (ii) suboptimality of the solution, and (iii) unavailability of efficient GPU utilization.

In this work, a fundamentally novel deep global routing approach is proposed. With this approach, a trained neural network (NN) receives pins' position and capacity per a global routing tile (i.e., the input features) and outputs a fully routed layout. This is done by mapping the global routing problem onto a two-dimensional image-to-image manipulation problem. The final image (i.e., the routed layout) is generated in a single step from the input image (i.e., the placed pins and tile capacity) variational autoencoder [5] (VAE) – a proven ML architecture for image processing problems. With NNs, the routing problem becomes naturally parallelizable on the existing GPU platforms. In this paper, a properly designed deep NN is demonstrated capable of efficiently learning routing patterns during the training process and detecting the patterns in inference, precisely determining the preferred routing heuristic. Three primary contributions of this work are, therefore:

1) Routing an unseen layout with a deep NN,
2) A unified, single-step non-iterative routing flow,
3) Parallelizing routing to systematically run on GPU hardware.

2 Problem formulation

Assume a typical global routing problem, in which the layout is split into rectangular tiles, the initial per-tile routing capacity is given, and numerous routing objectives need to be solved. A single-net routing objective is defined as a set of pins located at certain tiles and updated capacity of every tile in the layout. Routed net is defined by a set of adjacent tiles with a non-zero capacity that connect all the pins in a routing objective. Consider the following definitions for mapping a routing problem onto a supervised ML task, as illustrated in Figure 1.

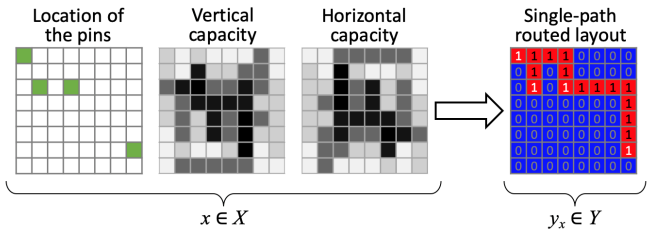


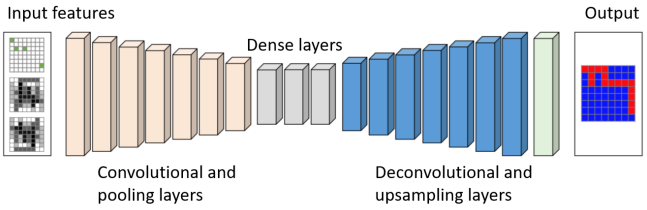
Figure 1: Single routing objective $x \in X$ and the corresponding single-net routed layout $y_x \in Y$. The tiles with maximum and zero routing capacity are, respectively, with white and black.

Let X and Y be sets of, respectively, single-net routing objectives and corresponding single-net routed layouts. A routing objective is to find the preferred routed path of tiles, $y_x \in Y$, connecting all the

978-1-7281-1085-1/20/$31.00 ©2020 IEEE

placed pins within the constraints of the vertical and horizontal per-tile capacity, as defined by $x \in X$. For any routing objective $x \in X$, the corresponding single-net routed layout $y_x \in Y$ is an $n \times n$ bitmap of tiles (i,j), $1 \le i, j \le n$, where each tile is associated with a binary score, $y_{i,j} = 0$ or $y_{i,j} = 1$ if the routing tile (i,j) is, respectively, excluded from or included within the preferred global routing path.

The goal is to train an ML system $f(X, \Theta)$ (where Θ is the trained model of the ML weights) that provides the conditional probability of each tile, $t(i,j)$, to be either included within (i.e., $f_{t_{(i,j)}} \ge 0.5$) or excluded from (i.e., $f_{t_{(i,j)}} < 0.5$) the preferred global routing, $f_{t(i,j)}(X, \Theta) = P(t_{(i,j)} = 1 | X)$. The proposed VAE architecture is shown in Figure 2 with detailed layers description.



	Layer type	filters	Kernel	Stride	Padding	Activation
1	Convolutional	32x32x32	3x3	1	Same	ReLu
2	Pool		2x2			
3	Convolutional	16x16x64	3x3	1	Same	ReLu
4	Pool		2x2			
5	Convolutional	8x8x128	3x3	1	Same	ReLu
6	Pool		2x2			
7	Convolutional	4x4x256			Same	ReLu
8	Hidden	256*				ReLu
9	Hidden	512*				ReLu
10	Hidden	256*				ReLu
11	Convolutional	4x4x256	3x3	1	Same	ReLu
12	Upsampling		2x2			
13	Convolutional	8x8x128	3x3	1	Same	ReLu
14	Upsampling		2x2			
15	Convolutional	16x16x64	3x3	1	Same	ReLu
16	Upsampling		2x2			
17	Convolutional	32x32x2		1	Same	Sigmoid

Figure 2: Structure of the proposed generative network.

Note that only a small portion of tiles (out of all the tiles) is included in each routing path, yielding an inherently unbalanced data set which tends to converge to the empty routing solution. To bias the VAE against the empty routes, a custom loss function is proposed to penalize the trained model if the number of tiles in the reference and ML paths are not the same: $f_L = \mathrm{MSE}(\hat{y}, y_{ref})(1 + k \cdot \mathrm{S} \cdot \mathrm{D})$, where $\mathrm{D} = \sum_{i,j} H(\hat{y}_{i,j}) - H(y^{\mathrm{ref}}_{i,j})$, $\mathrm{S} = k_{err} \cdot \mathrm{sign}(D - 1) + 1$. Here H(·) is the Heaviside step function and MSE is the mean squared error. In this paper, the proposed loss function is used with $k = 10^{-3}$ and $k_{err} = 10^2$.

The training data set $\{(X_i, Y_i)\}_{i=1}^{N}$ comprises N synthetic routing objectives in the bitmap representation and N corresponding reference single-net routed layouts (i.e., the true labels). To generate this training set, hundreds of thousands of reference routes are generated and routed with the deterministic FastRoute router. VAE model is trained using dynamic moment and learning rate management algorithm RMSProp for a thousand of epochs [6]. The model produces meaningful, but incorrect outputs after a few minutes of training, converging to a legit and effective model after hours of training, as shown in Figure 3. The properly trained ML router autonomously finds optimal heuristics, efficiently routing unseen layouts without retraining.

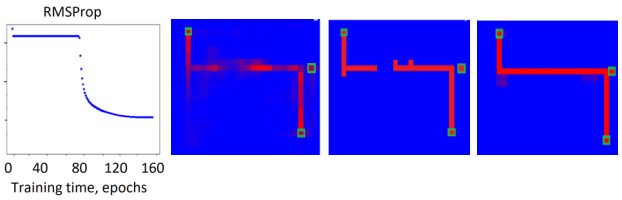


Figure 3: From left to right: a training loss function and the output of two unsaturated models and a fully saturated model.

3 Evaluation

The VAE based router is evaluated based on all the two- and three-pin nets (a total of 8,677 previously unseen nets) of the ISPD'98 ibm01 64x64 circuit. The VAE based router and the state-of-the-art FastRoute are both executed on this test set on the AMD Ryzen CPU 3.2GHz workstation with NVIDIA GTX1080 GPU support. The runtime and routability (the per cent of the successfully routed nets out of the total attempted nets) with the VAE based router are, respectively, 18.5% and 96.8%, as compared to the FastRoute performance on the same test set.

4 Conclusions and Future Work

Fundamentally new global routing approach is demonstrated in this paper. Unlike all the existing methods, this approach treats the global routing problem as an image processing problem and solves it with a deep learning system. As a result, the router can be efficiently executed on a typical GPU, yielding over five times shorter runtime with similar to state-of-the-art routability. As a proof of concept, the proposed solution has two primary limitations: 1) the training runtime increases with the resolution of the layout, and 2) the performance of the proposed solution is limited by the performance of the reference router. To address the first concern, the routing training sets should be accumulated and enhanced over generations of new designs and technologies. In addition, unsupervised and semi-supervised learning is currently being explored by the authors to reduce the training complexity and enhance the inference performance, as currently limited by a reference router.

REFERENCES

[1] I. L. Markov, et al., "Progress and Challenges in VLSI Placement Research," *Proc. of IEEE*, 103(11), pp. 1985-2003, 2015.

[2] Z. Qi, et al., "Accurate Prediction of Detailed Routing Congestion Using Supervised Data Learning," *IEEE Int. Conf. on Comp. Design*, 2014, pp. 97-103.

[3] S. M. Sait and Umair F. Siddiqi, "A Stochastic Evolution Algorithm Based 2D VLSI Global Router," *Integration*, vol. 53, pp. 125-125, 2016.

[4] Liao, Haiguang, et al. "A Deep Reinforcement Learning Approach for Global Routing," *J. of Mechanical Design*, 142(6), 2020.

[5] P. Isola, et al., "Image-to-Image Translation with Conditional Adversarial Networks," *Proc. of the IEEE Conf. on CVPR*, 2017.

[6] Y. N. Dauphin, et al., "Equilibrated Adaptive Learning Rates for Non-Convex Optimization", *Proc. of the Int. Conf. on NIPS*, vol. 1, pp 1504–1512, 2015.