

DOI:10.19651/j.cnki.emt.2415567

# 基于 ResCSP-34 的集成电路供电网络静态电压降预测研究 \*

李 岳 夏益民

(广东工业大学集成电路学院 广州 510006)

**摘要:** 随着超大规模集成电路的不断发展,片上供电网络的设计日益重要,需要通过计算供电网络的静态电压降来反映设计的性能。然而传统的计算方法需要耗费大量的时间,导致芯片设计周期延长。为了缩短芯片设计的周期和提高芯片设计的效率,本文提出了一个基于卷积神经网络—ResCSP-34 的快速静态电压降预测模型。模型采用编码器-解码器结构,首先对残差网络 ResNet34 进行修改作为编码器的主体结构,然后在解码器中引入特征融合模块,并且在编码器和解码器的连接处引入注意力机制模块,最后提出了一个同时结合了均方误差、皮尔逊相关系数和平均绝对误差的损失函数对模型进行训练。实验结果表明,在 CircuitNet 数据集上,模型预测结果的平均绝对误差为 0.7 mV,小于 1 mV,皮尔逊相关系数的平均值大于 0.93,接近于 1,对一个片上供电网络设计进行静态电压降预测的平均总时间为 7.36 s,其中卷积神经网络的平均推理时间为 0.015 s。实验结果表明,ResCSP-34 模型能够快速且精准地预测静态电压降。

**关键词:** 集成电路;静态电压降;卷积神经网络;编码器-解码器;注意力机制

中图分类号: TN791 文献标识码: A 国家标准学科分类代码: 510

## Research on static voltage drop prediction of IC power supply network based on ResCSP-34

Li Yue Xia Yimin

(School of Integrated Circuits, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract:** With the continuous development of VLSI circuits, the design of the on-chip power delivery network is becoming increasingly important, and the performance of the design needs to be reflected by calculating the quiescent voltage drop of the power delivery network. However, traditional computational methods are time-consuming, resulting in longer chip design cycles. In order to shorten the cycle of chip design and improve the efficiency of chip design, this paper proposes a fast static voltage drop prediction model based on convolutional neural network—ResCSP-34. The model adopts the encoder-decoder structure, firstly the residual network ResNet34 is modified as the main structure of the encoder, then the feature fusion module is introduced into the decoder, and the attention mechanism module is introduced at the connection of the encoder and the decoder, and finally a loss function combining the mean square error, Pearson correlation coefficient and mean absolute error is proposed to train the model. Experimental results show that on the CircuitNet dataset, the average absolute error of the model prediction results is 0.7 mV, which is less than 1 mV, the average value of the Pearson correlation coefficient is greater than 0.93, close to 1, and the average total time for static voltage drop prediction for an on-chip power supply network design is 7.36 s, and the average inference time of the convolutional neural network is 0.015 s. Experimental results show that the ResCSP-34 model can quickly and accurately predict the quiescent voltage drop.

**Keywords:** integrated circuits; static voltage drop; convolutional neural networks; encoder-decoder; attention mechanism

## 0 引言

VLSI 和制造工艺节点的不断发展,电子设计自动化 (electronic design automation, EDA) 工具在芯片设计的流程中越来越重要和不可或缺<sup>[1]</sup>。尤其是在后端的物理设计

收稿日期:2024-03-02

\* 基金项目:广东省重点领域研发计划(2022B0701180001)项目资助

中,使用 EDA 工具对物理设计的相关指标进行有效的预测,有助于提前规避在物理设计后期可能会出现的问题,缩短物理设计的迭代周期,减少重新设计所需的时间,提高设计效率<sup>[2-3]</sup>。

片上供电网络(power delivery network, PDN)的设计是物理设计中十分重要的一步,影响着芯片的功耗、性能和面积<sup>[4]</sup>,设计人员通常需要花费大量的时间进行仿真,通过分析 PDN 的电压降(IR Drop)来反映 PDN 的设计性能。传统的商业 EDA 工具一般通过求解具有数百万到数十亿个变量的大型方程组得到 IR Drop,这需要耗费大量时间和硬件计算资源。同时,多次的迭代设计和仿真会延长芯片设计周期,增加设计成本,从而迫切需要一个方法来快速且精准地得到 IR Drop。

近年来随着人工智能的飞速发展,人工智能算法已经在许多领域中取得了令人瞩目的成绩<sup>[5]</sup>。在 EDA 领域中,已经有很多研究人员将人工智能算法应用到物理设计中,极大地提高了物理设计的效率。如 Lu 等<sup>[6]</sup>将图神经网络(graph neural network, GNN)和长短记忆(long short term memory, LSTM)神经网络相结合,能够帮助设计人员在物理设计早期评估当前设计的性能,可以极大地提高设计效率。Xie 等<sup>[7]</sup>在布局阶段使用卷积神经网络来评估布局方案的可布线性,能够在众多的布局设计方案中快速地筛选出合适的方案。目前将人工智能算法和 IR Drop 预测相结合的研究比较少,文献[8]提出一个基于卷积神经网络的预测模型—PowerNet,但是需要将一个完整的版图切割为多个方块,然后依次传入到模型中进行预测,导致模型的预测速度相对较慢且没有考虑版图的整体信息。Chhabria 等<sup>[9]</sup>提出了基于语义分割模型 UNet 的预测模型,将一个完整的版图信息映射到一个低分辨率的图上,直接传入到模型中进行预测,实现了快速的端到端的预测,但是精度不高。

编码器-解码器结构是神经网络中一个十分重要的模型结构,在处理各种复杂的任务中均有优异的性能表现。在卷积神经网络中,UNet<sup>[10]</sup>、SegNet<sup>[11]</sup> 和 Deeplab<sup>[12-14]</sup> 等经典的模型均采用编码器-解码器结构。这些模型的编码器通常是 VGG<sup>[15]</sup>、残差网络 ResNet<sup>[6]</sup> 等特征提取网络,负责对输入的数据进行特征提取,捕获输入数据的关键特征信息,然后通过跳跃连接在解码器中进行上采样、特征融合等操作实现端到端的输出。

注意力机制被广泛地应用于神经网络中以提高网络的性能,在编码器和解码器的连接通路中引入注意力机制,可以提高编码器和解码器之间的信息传递,使得解码器可以关注到编码器提取的特征信息中更加重要的部分,从而更加有效地利用编码器的特征信息生成输出。

为了可以实现快速且精准的静态电压降预测,本文的主要工作如下:

1) 提出了一个基于编码器-解码器结构的卷积神经网

络 ResCSP-34 用于静态电压降的预测。ResCSP-34 的编码器部分采用调整后的残差网络 ResNet34 作为主体结构进行特征提取,在解码器中通过转置卷积和特征融合模块 CSP(cross stage partial)<sup>[17]</sup> 实现上采样和特征融合,并且在编码器和解码器的跳跃连接通路中引入 CBAM(convolutional block attention module)<sup>[18]</sup> 注意力机制模块进一步提高模型的性能。使用 ResCSP-34 进行静态电压降预测可以充分地考虑版图的全局信息,以物理设计流程中的相关数据作为输入,输出高精度的 IR Drop 数据。

2) 提出了一个同时将均方误差(mean squared error, MSE)、皮尔逊相关系数(pearson correlation coefficient, PCC)和平均绝对误差(mean absolute error, MAE)作为损失项的损失函数,使用该损失函数对模型进行训练可以进一步提高模型的预测性能。

## 1 静态电压降的预测

电压降主要分为静态电压降和动态电压降。本文将对静态的电压降预测进行研究,在后面的表述中使用 IR Drop 来统一表示静态的电压降。传统的计算方法首先将供电网络抽象为线性的数学模型,然后通过求解矩阵方程来计算静态电压降<sup>[19]</sup>,矩阵方程如式(1)所示。

$$\mathbf{Ax} = \mathbf{b} \quad (1)$$

其中,矩阵  $\mathbf{A}$  是供电网络的电导矩阵,向量  $\mathbf{x}$  是节点的电压向量,  $\mathbf{b}$  是电源向量,矩阵  $\mathbf{A}$  的计算公式如式(2)所示。

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{D}_1 & \cdots & \cdots \\ \mathbf{D}_1 & \mathbf{A}_{22} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \mathbf{D}_{N-1} \\ \cdots & \cdots & \mathbf{D}_{N-1} & \mathbf{A}_{NN} \end{bmatrix} \quad (2)$$

其中,  $N$  表示供电网络模型的节点数,  $\mathbf{A}_{ii}$  和  $\mathbf{D}_i$  的计算公式如式(3)、(4)所示。

$$\mathbf{A}_{ii} = \begin{bmatrix} G_{(i,1)} & -g_{(i,2)(i,1)} & \cdots & \cdots \\ -g_{(i,2)(i,1)} & G_{(i,2)} & \cdots & \cdots \\ \ddots & \ddots & \ddots & \ddots \\ \cdots & \cdots & -g_{(i,N)(i,N-1)} & G_{(i,N)} \end{bmatrix} \quad (3)$$

$$\mathbf{D}_i = \begin{bmatrix} -g_{(i,1)(i+1,1)} & 0 & \cdots & 0 \\ 0 & -g_{(i,1)(i+1,1)} & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & \cdots & -g_{(i,1)(i+1,N)} \end{bmatrix} \quad (4)$$

其中,  $g_{(i,j)(k,l)}$  表示节点  $(i,j)$  和节点  $(k,l)$  之间的电导,  $G_{(i,j)}$  的计算如式(5)所示。

$$G_{(i,j)} = \sum_{(k,l) \in N_{(i,j)}} |g_{(i,j)(k,l)}| \quad (5)$$

其中,  $(k,l) \in (i,j)$  表示节点  $(i,j)$  周围的节点。向量  $\mathbf{b}$  的计算公式如式(6)所示。

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_N \end{bmatrix} \quad \mathbf{b}_i = \begin{bmatrix} -I_{(i,1)} \\ -I_{(i,2)} \\ \vdots \\ -I_{(i,N)} \end{bmatrix} \quad (6)$$

其中,  $I_{(i,j)}$  表示节点  $(i,j)$  的电流。

从式(2)和(3)可以看出,矩阵  $\mathbf{A}$  是一个  $N^2 \times N^2$  的矩阵,对于一个 VLSI 来说一般具有上百万个节点数  $N$ ,导致方程的求解通常需要几十到上百个小时,这往往是物理验证中最消耗资源和时间的仿真环节。

目前使用深度学习来进行 IR Drop 预测的方法可以分为以下两类:

1)首先将整个物理版图划分为多个方块,然后将这些方块依次传入到模型中进行预测,最后经过多次预测得到最终的预测结果。因为需要多次进行模型推理,所以预测速度相对较慢且不能充分考虑全局信息。

2)将整个物理版图作为整体,传入到模型中进行预测,只需要经过模型一次就可以得到预测结果。

本文所提出的预测模型属于第二种,可以快速地进行 IR Drop 预测。

## 2 基于 ResCSP-34 的静态电压降预测

### 2.1 整体方案

本文提出的整体技术方案如图 1 所示,首先对相关报告进行特征提取,得到训练集,然后对训练集中数量较少的数据进行扩充,最后使用扩充后的训练集对 ResCSP-34 进行训练。

在训练过程中,首先将图像数据进行组合得到对应的特征图和标签图,然后将特征图传入到 ResCSP-34 模型中进行预测,得到对应的预测图,接着计算预测图和标签图之间的损失值,判断损失值是否收敛,当损失值没有收敛时,反向更新模型的权重,当损失值收敛时结束训练。

测试时同样地对测试的报告进行特征提取,然后组合成特征图传入到 ResCSP-34 中进行预测,得到预测图,最后转化为预测报告。

### 2.2 特征提取

在物理设计中,会产生一系列报告文件,如等效电阻文件,功率文件和电压降文件等。为了可以使用卷积神经网络来进行 IR Drop 预测,需要先对报告文件中的数据进行特征提取,将报告文件中的数据映射到图像中,得到相对应的特征图,然后再传入到模型中进行预测。在这些报告中,包含了每一个器件的中心点坐标以及对应的数据,如功率数据、等效电阻数据或电压降等。首先根据版图的大小创建一个像素值全为 0 的图像,然后将器件的中心点坐标作为映射到图像中的像素点坐标,像素值就等于对应的功率数据或等效电阻数据,最后再将图像的大小尺寸统一转化为  $256 \times 256$ 。

影响 IR Drop 大小的因素有很多,其中器件的自身总

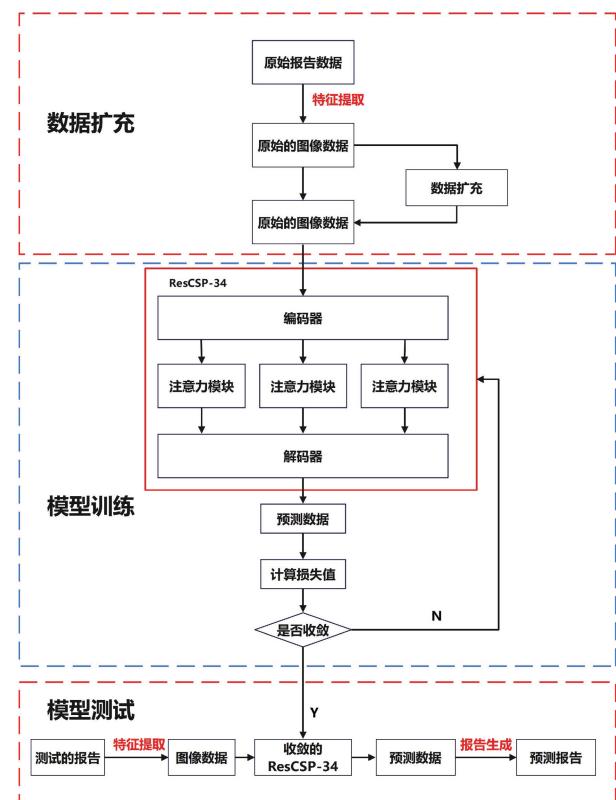
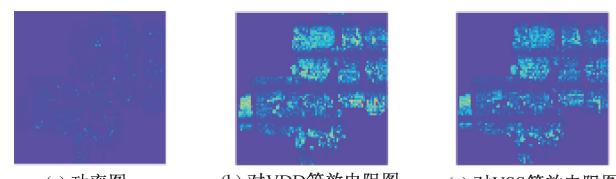


图 1 整体技术方案

功率和到电源网络的等效电阻影响最大,所以选择版图对应的功率图、到电源 VDD 的等效电阻图和到电源地 VSS 的等效电阻图作为特征图,如图 2 所示。将这 3 个特征图在通道维度上进行拼接,然后传入到模型中进行预测,得到对应的预测图。



(a) 功率图 (b) 对VDD等效电阻图 (c) 对VSS等效电阻图

图 2 特征图可视化

IR Drop 等于对电源 VDD 的压降(vdd drop)和对电源地 VSS 的压降(Gnd Bounce)之和,为了进一步提高预测的准确性,将 VDD\_Drop 和 Gnd Bounce 作为标签进行训练,VDD\_Drop、Gnd Bounce 和 IR Drop 的可视化如图 3 所示。

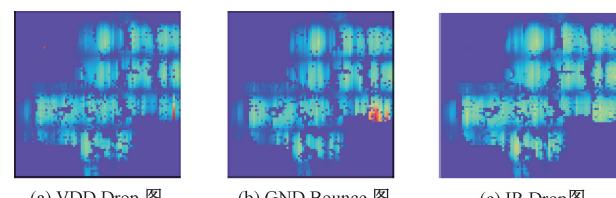


图 3 标签图可视化

### 2.3 数据扩充

在本文使用的数据集中存在 2 个类型的设计数据量极少,如表 1 所示。为了使模型具有泛化性,需要对这两类设计的数据进行适当的扩充。由于数据量少,无法使用一些生成式模型进行数据扩充,如对抗式生成网络、扩散模型等,所以本文使用传统的数据增强方法对这两类数据进行扩充,具体的流程如图 4 所示。根据生成的随机数范围,选择对应的方法进行数据扩充,在对特征图进行操作时,与其对应的标签图也需要进行对应的操作,确保训练数据的有效性。

### 2.4 ResCSP-34 模型结构

为了可以实现端到端的 IR Drop 预测,本文提出了一个卷积神经网络—ResCSP-34,如图 5 所示。ResCSP-34 采用编码器-解码器结构,编码器部分采用经过修改的 ResNet34 作为主体结构,解码器部分进行上采样恢复特征图的大小,并且添加了 CSP 模块,使得解码器的特征融合效果更好,进一步提高模型的整体性能。

在连接编码器和解码器的通路中,加入了 CBAM 注意力机制模块,使得模型自适应学习编码器输出特征图的重要程度,得到更加高质量的特征图,从而使得解码器最后输出的预测图更加符合实际的 IR Drop 图。

#### 1) 编码器结构

因为残差网络 ResNet 能够获得更好的特征表示,被广

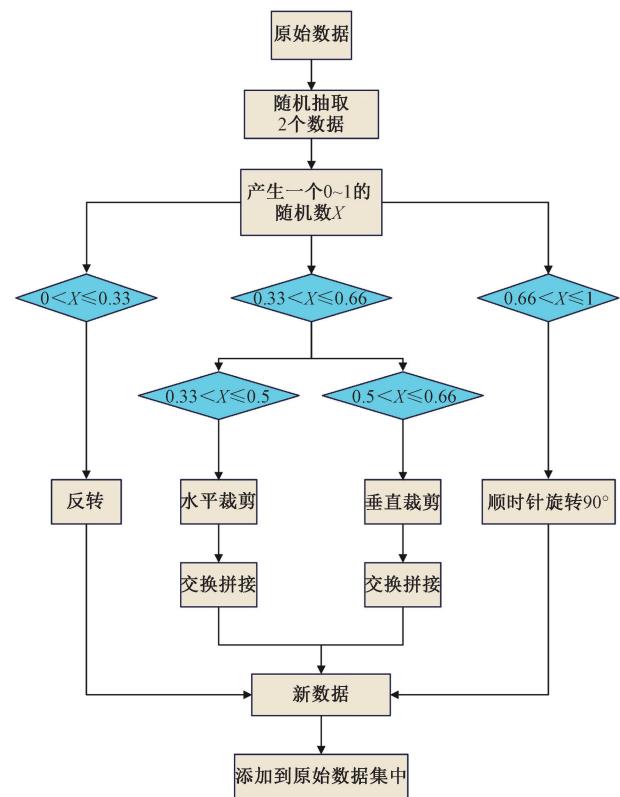


图 4 数据扩充过程

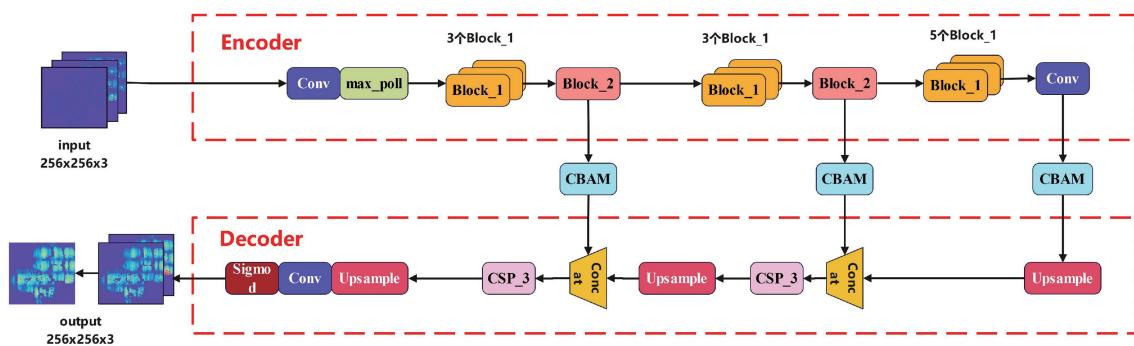


图 5 ResCSP-34 的结构图

泛用于特征提取,所以综合考虑 IR Drop 预测任务的复杂程度,选择对 ResNet34 进行修改作为 ResCSP-34 编码器的主体结构。使用尺寸为  $2 \times 2$ ,步长为 2 的最大池化替换第一层的  $7 \times 7$  卷积下采样并且为了不让特征图的尺寸过小,导致信息损失过多,删除 ResNet34 的最后几个残差块。残差块 Block\_1 和 Block\_2 的结构如图 6 所示,其中 CIL 是一个卷积层,结构如图 7 所示,Block\_1 不改变特征图的通道数和尺寸大小,Block\_2 对特征图进行 2 倍的下采样和升通道数。

为了弥补下采样过程中导致的信息损失,在编码的过程中,每经过一系列的 Block\_1 后都会输出一个特征图,从图 5 中可以看出编码器最终会输出 3 个不同尺寸的特征

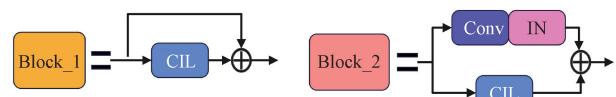


图 6 Block\_1 和 Block\_2 的结构图

$$\text{CIL} = \text{Conv} \quad \text{IN} \quad \text{Leaky relu} \quad \text{Conv} \quad \text{IN} \quad \text{Leaky relu}$$

图 7 CIL 模块结构

图,这 3 个特征图会在后面的解码器中进行特征融合,使得最终得到的预测图精度更高。

#### 2) 注意力机制

在卷积神经网络中常用的注意力模块有 SENet<sup>[20]</sup>、

ECANet<sup>[21]</sup>、CBAM 和 BAM<sup>[22]</sup>等。本文将采用 CBAM 注意力模块对编码器输出的特征图进行注意力加权。CBAM 模块是一个同时结合了通道注意力机制和空间注意力机制的模块,其结构如图 8 所示。

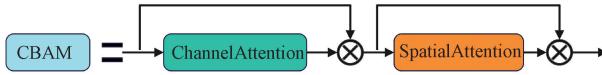


图 8 CBAM 模块的结构图

CBAM 的计算公式如式(7)、(8)所示,特征图  $f$  首先经过一个通道注意力模块得到特征图  $f_c$ ,然后  $f_c$  再经过一个空间注意力模块输出特征图  $f_{cs}$ 。

$$f_c = \text{Channel}(f) \quad (7)$$

$$f_{cs} = \text{Spatial}(f_c) \quad (8)$$

通道注意力模块的结构如图 9 所示,计算公式如式(9)~(11)所示,首先在通道维度上通过最大池化和平均池化计算得到权重  $W_{max}$  和  $W_{avg}$ ,然后将两者相加再经过  $Sigmoid$  函数得到最终的通道权重,最后将通道权重和  $f$  相乘得到  $f_c$ 。

$$W_{max} = \text{Conv}(\text{Relu}(\text{Conv}(\text{maxpoll}(f)))) \quad (9)$$

$$W_{avg} = \text{Conv}(\text{Relu}(\text{Conv}(\text{avgpoll}(f)))) \quad (10)$$

$$\text{Channel}(f) = \text{Sigmoid}(W_{max} + W_{avg}) \times f \quad (11)$$

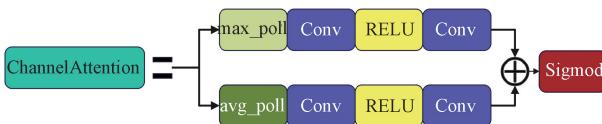


图 9 通道注意力模块

空间注意力模块的结构如图 10 所示,计算公式如式(12)、(13)所示,首先将在空间维度上进行最大池化和平均池化的结果在通道维度上进行拼接,然后再经过一个卷积和  $Sigmoid$  函数得到空间权重,最后将空间权重和  $f_c$  相乘输出  $f_{cs}$ 。

$$W_{max\_avg} = \text{Conv}(\text{Concat}(\text{maxpoll}(f_c) + \text{avgpoll}(f_c))) \quad (12)$$

$$\text{Spatial}(f_c) = \text{Sigmoid}(W_{max\_avg}) \times f_c \quad (13)$$

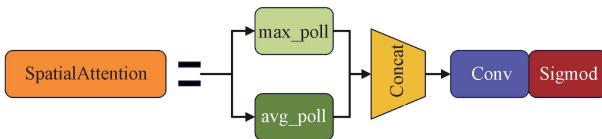


图 10 空间注意力模块

### 3)解码器结构

ResCSP-34 的解码器主要对特征图进行上采样和特征融合,输出预测图。上采样 Upsample 模块的结构如图 11 所示,主要通过转置卷积 ConvTrans 实现。经过 3 次上采样后,得到的预测图大小和编码器输入的特征图大小一样,从而实现端到端的预测。

从图 5 可以看出,前面的 2 次上采样后都会和编码器



图 11 上采样模块

输出的另外一个特征图在通道维度上进行拼接,然后经过一个 CSP\_X 模块进行特征融合,CSP\_X 模块结构如图 12 所示。

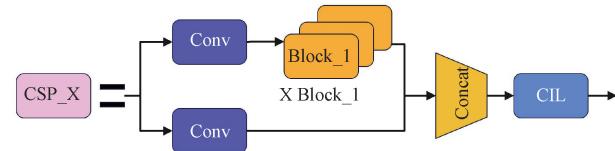


图 12 CSP\_X 结构图

CSP\_X 本质上是一个残差模块,可以分为 2 个通路,一个是主通路,由 1 个  $1 \times 1$  的卷积和 X 个残差块 Block\_1 组成,另一个是残差边通路,是 1 个  $1 \times 1$  的卷积,最后将 2 个通路的结果在通道维度上进行拼接,再经过一个 CIL 模块得到输出。

在经过一系列的模块后,解码器会输出 Vdd Drop 预测图和 Gnd Bounce 预测图,将这两个预测图进行相加得到最后的 IR Drop 预测图,后续根据这个预测图生成预测报告。

### 2.5 损失函数

在训练语义分割模型时,因为模型是对像素点类别进行预测,所以一般用交叉熵损失函数进行训练,具体表达式如式(14)所示。

$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K C_{ij} \times \ln C'_{ij} \quad (14)$$

其中,  $N$  为像素点的总数,  $K$  为类别的总数,当像素点  $i$  的类别为  $j$  时,  $C_{ij}$  等于 1,否则等于 0,  $C'_{ij}$  为预测像素点  $i$  属于种类  $j$  的概率。

但是 IR Drop 预测任务并不是一个像素点类别预测的任务,而是预测像素点具体的像素值,所以不适合使用一般的交叉熵损失函数进行训练。均方误差 MSE 和平均绝对误差 MAE 可以用来计算图像间像素值的差异。然而,由于 MSE 损失函数是平方误差计算,这将导致其对异常值十分敏感,即使只有一个异常值,也有可能导致损失值大幅度增加,从而影响模型的性能。此外,MAE 损失函数并非处处可导,且赋予所有误差相同的梯度,这导致模型优化过程并不光滑,模型难以收敛。

为了使模型的训练效果更好,本文提出了一个新的损失函数对 ResCSP-34 进行训练,函数表达式如式(15)所示。

$$\text{Loss} = W_1 \times \alpha \times \text{MSE} + W_2 \times \beta \times (1 - \text{PCC}) \times 0.001 + W_3 \times \lambda \times \text{MAE} \quad (15)$$

该函数同时结合了均方误差 MSE、相关系数误差 1-PCC 和平均绝对误差 MAE 作为损失项,其中  $W_1$ 、 $W_2$  和

$W_3$  分别是 3 个损失项对应的权重, 分别设置为  $W_1 = 0.4$ 、 $W_2 = 0.2$  和  $W_3 = 0.4$ ,  $\alpha$ 、 $\beta$  和  $\lambda$  分别是 3 个损失项对应的缩放因子, 分别设置为  $\alpha = 100$ 、 $\beta = 0.001$  和  $\lambda = 1$ 。

1) MSE 损失项和 MAE 损失项, 衡量预测图和标签图之间像素值的差异, 计算公式如式(16)和(17)所示。

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad (16)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_i - \hat{Y}_i| \quad (17)$$

其中,  $N$  为像素点的总数,  $Y$  为标签图的像素值,  $\hat{Y}$  为预测图的像素值。

2)(1-PCC)损失项, 衡量预测图和标签图之间线性关系的不相似性, 皮尔逊相关系数 PCC 的计算公式如式(18)所示。

$$PCC = \frac{\sum_{i=1}^N [Y_i - mean(Y_i)][\hat{Y}_i - mean(\hat{Y}_i)]}{\sqrt{\sum_{i=1}^N [Y_i - mean(Y_i)]^2 \sum_{i=1}^N [\hat{Y}_i - mean(\hat{Y}_i)]^2}} \quad (18)$$

其中,  $N$  为像素点的总数,  $Y$  为标签图的像素值,  $\hat{Y}$  为预测图的像素值。

### 3 实验与结果分析

#### 3.1 实验环境

硬件配置: GPU 为 NVIDIA GA100 40 GB; CPU 为 Intel(R) Xeon(R) Platinum 8368 CPU@2.40 GHz 152; 内存 1 T。

软件配置: Ubuntu20.04.6 操作系统; Python3.9; Pytorch2.10。

#### 3.2 实验数据

本文使用 CircuitNet<sup>[23]</sup> 数据集中 5 个类型的数据作为实验数据, 该数据使用 5 个原始的 RTL 设计作为基准电路, 然后通过设置不同的逻辑综合参数和物理后端的参数得到 7 321 个数据作为实验数据, 具体的设计类型和对应的数据量如表 1 所示。从表 1 中可以看出, 对于 nvnda-small 和 Vortex-small 这两个类型的设计, 原始数据量极少, 所以为了提高模型的泛化性, 先使用 2.3 节中数据扩充方法对这两个类型的数据分别扩充到 1 000 个, 然后随机对这

些数据进行划分, 其中 80% 作为训练集, 另外 20% 作为测试集。

表 1 数据集中各种类型的数量

设计类型	数量/个
RISCY-FPU	2 400
RISCY	2 400
zero-riscy	2 400
nvdla-small	55
Vortex-small	66

#### 3.3 实验结果和分析对比

##### 1) 不同模型间的比较

实验中使用 2 个类型的指标来评估模型的性能, 分别是平均绝对误差 MAE 和皮尔逊相关系数 PCC, 具体的计算公式如式(17)、(18)所示, MAE 越小越好, PCC 越接近 1 越好。

将本文设计的 ResCSP-34 和同样可以实现端到端输出的模型 FCN<sup>[24]</sup>、UNet、SegNet、ConvNeXt<sup>[25]</sup> 和 ViT (vision transformer)<sup>[26]</sup> 进行对比, 每个模型在相同的超参数下进行训练, 对应的超参数如表 2 所示, 得到的测试结果如表 3 所示, 可视化结果如图 13 所示, 测试结果的箱型图如图 14 所示。

表 2 训练超参数

参数	值
Batch_Size	10
学习率	0.002
迭代次数	30

表 3 不同模型的测试结果

模型	MAE/v	PCC
FCN	0.002 8	0.753 0
UNet	0.003 8	0.471 9
SegNet	0.003 0	0.445 3
ConvNeXt	0.004 9	-0.000 7
ViT	0.004 4	-0.079 5
ResCSP-34	<b>0.001 1</b>	<b>0.881 9</b>

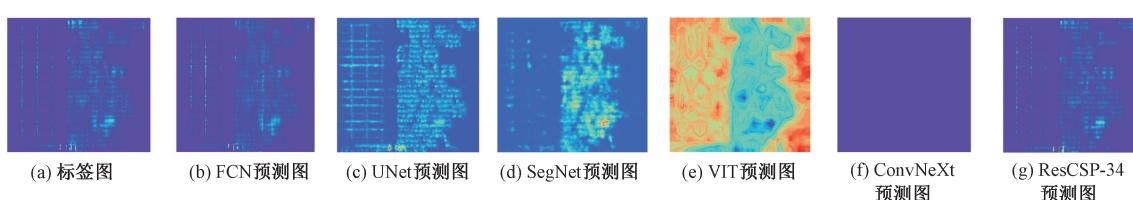


图 13 不同模型测试结果

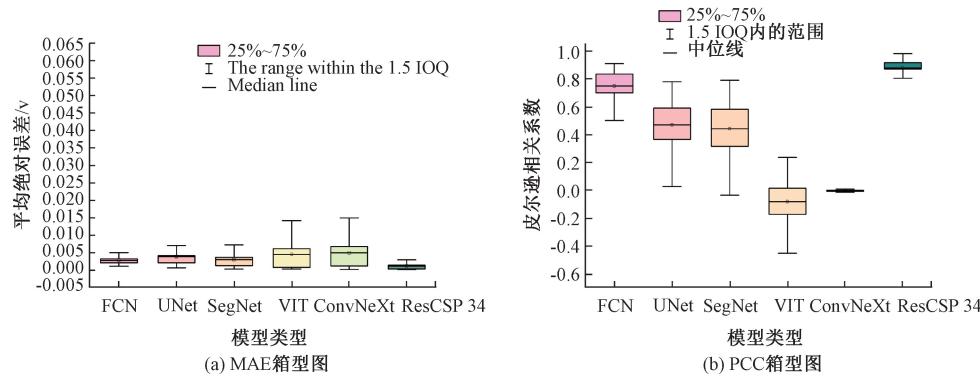


图 14 测试结果的箱型图

从上面的测试数据可以看出,本文所提出的 ResCSP-34 在相同的训练参数下,MAE 和 PCC 这两个指标均为最优。MAE 比 FCN 低 60.7%,比 UNet 低 71%,比 SegNet 低 63.3%,比 ConNext 低 77.6%,比 VIT 低 75%。相关系数 PCC 比 FCN 高 17.1%,比 UNet 高 86.8%,比 SegNet 高 98.1%。ConNext 和 VIT 的相关系数为负数,表明不适合采用这两个模型预测 IR Drop。从测试的箱型图可以更加直观地看出 ResCSP-34 较其他模型稳定性更好。

### 2) 不同损失函数间的比较

分别将式(19)~(21)作为损失函数对 ResCSP-34 进行训练,使用的超参数如表 2 所示,式(19)只考虑了均方误差 MSE 作为损失项,式(20)同时考虑了均方误差 MSE 和相关系数 PCC 作为损失项,式(21)则同时考虑了均方误差 MSE、相关系数 PCC 和平均绝对误差 MAE 作为损失项,得到的测试结果如表 4 所示。

$$\text{Loss1} = \text{MSE} \times 100 \quad (19)$$

$$\text{Loss2} = 0.5 \times \text{MSE} \times 100 + 0.5 \times (1 - \text{PCC}) \times 0.001 \quad (20)$$

$$\text{Loss3} = 0.4 \times \text{MSE} \times 100 + 0.2 \times (1 - \text{PCC}) \times 0.001 + 0.4 \times \text{MAE} \quad (21)$$

表 4 ResCSP-34 模型使用不同损失函数的测试结果

损失函数	MAE/v	PCC
Loss1	0.001 6	0.812 1
Loss2	0.001 3	0.852 3
<b>Loss3</b>	<b>0.001 1</b>	<b>0.881 9</b>

从实验结果可以看出,使用 Loss3 对模型进行训练时,训练效果最好,MAE 比 Loss1 低 31.3%,比 Loss2 低 15.4%,PCC 比 Loss1 高 8.6%,比 Loss2 高 3.5%。

### 3) 模型最优的结果

通过调整模型训练的迭代次数,发现模型迭代到 160 多次时就趋于收敛,模型迭代训练 160 次时 MAE 和 PCC 的测试结果如表 5 所示。

表 5 ResCSP-34 迭代训练 160 次 MAE 和 PCC 的测试结果

设计类型	数量 / 个	ResCSP-34		ResCSP-34(扩充)	
		MAE/v	PCC	MAE/v	PCC
RISCY-FPU	682	0.000 7	0.948 4	0.000 7	0.936 1
RISCY	670	0.000 7	0.942 1	0.000 8	0.929 7
zero-riscy	733	0.000 6	0.941 7	0.000 7	0.930 0
nvdla-small	20	0.002 1	0.764 3	0.001 8	0.793 1
Vortex-small	21	0.001 0	0.868 6	0.000 9	0.890 9
Average	—	0.000 7	0.941 6	0.000 7	0.930 2

从 MAE 和 PCC 的测试结果可以看出,数据扩充前 ResCSP-34 在 5 类设计的预测中,平均 MAE 小于 1 mV,只有 0.7 mV,相关系数为 0.941 6,数据扩充后 ResCSP-34 对 nvdla-small 和 Vortex-small 这两类设计的平均 MAE 分别下降了 14.2% 和 10%,平均 PCC 分别提高了 3.7% 和 2.5%,而整体的平均 MAE 和平均 PCC 变化不大,分别为 0.7 mV 和 0.930 2。在进行了数据扩充后,模型的泛化性能够进一步提高。

进行一次 IR Drop 预测的全流程时间如表 6 所示。从表 6 可以看出,模型的平均推理时间仅有 0.015 s,而在一个数字 IC 中,通常具有几十万到上百万个器件,所以特征提取和最后的报告生成所耗费的时间最多,分别是 6.53 s 和 0.82 s,平均总时间为 7.36 s,与需要十几个小时的传统计算方法相比,本文所提出的方法在速度上具有明显的优势。

表 6 一次 IR Drop 预测的时间

过程	平均时间/s
特征提取	6.53
模型推理	0.015
报告生成	0.82
总时间	7.36

## 4 结论

本文提出了一个基于卷积神经网络—ResCSP-34 的

预测模型,首先对相关的报告文件进行特征提取,得到相对应的特征图,然后传入到 ResCSP-34 中进行预测,得到预测图,最后生成相对应的 IR Drop 预测报告。其中 ResCSP-34 基于编码器-解码器结构,并且在编码器和解码器的连接处引入了 CBAM 注意力机制模块进一步提高模型的性能。实验结果表明,和其他模型相比,在相同的超参数训练情况下,ResCSP-34 预测结果的 MAE 和 PCC 均为最优。为了使得模型的训练效果更好,本文设计了一个同时考虑了均方误差 MSE、皮尔逊相关系数 PCC 和平均绝对误差 MAE 的损失函数来对模型进行训练,实验结果表明,使用同时考虑了 MSE、PCC 和 MAE 的损失函数进行训练时,模型的性能更好。最后通过调整训练的迭代次数使得 ResCSP-34 趋于收敛,预测结果的平均 MAE 在 1 mV 以内,平均相关系数在 0.93 以上,并且一次 IR Drop 预测的平均总时间只有 7.36 s,和传统的预测方法相比,具有准度高、速度快的优点,能够提高设计人员进行 IR Drop 分析的速度,从而提高物理设计的效率。由于本文提出的预测模型输入输出大小固定,所以在进行尺寸变换的过程中会导致额外的误差,在未来的研究中可以进一步对模型的结构进行改进,使得模型能够自适应输入输出的大小,从而减少额外的误差。

## 参考文献

- [1] YU B, XU X, ROY S. Design for manufacturability and reliability in extreme-scaling VLSI [J]. Science China(Information Sciences), 2016, 59(6): 96-118.
- [2] YU C, HUANG C C, NAM G J, et al. End-to-end industrial study of retiming[C]. 2018 IEEE Computer Society Annual Symposium on VLSI(ISVLSI), IEEE, 2018: 203-208.
- [3] 王佳伟,黄志洪,高同强,等.一种高精度的 FPGA 电路面积时序预测方法[J].国外电子测量技术,2017, 36(6):31-35.
- [4] HO C T, KAHNG A B. IncPIRD: Fast learning-based prediction of incremental IR Drop [C]. 2019 IEEE/ACM International Conference on Computer-Aided Design(ICCAD), IEEE, 2019:1-8.
- [5] 聂廷远,徐坤鹏,孔琪.基于注意力机制的 FPGA 布线拥塞预测[J].电子测量技术,2023,46(11):159-165.
- [6] LU Y C, NATH S, KHANDEWAL V, et al. Doomed run prediction in physical design by exploiting sequential flow and graph learning[C]. 2021 IEEE/ACM International Conference on Computer Aided Design(ICCAD). Munich, Germany, 2021: 1-9.
- [7] XIE Z, HUANG Y H, FANG G Q, et al. RouteNet: Routability prediction for mixed-size designs using convolutional neural network[C]. 2018 IEEE/ACM International Conference on Computer-Aided Design(ICCAD), IEEE, 2018: 1-8.
- [8] XIE Z, REN H, KHAILANY B, et al. PowerNet: Transferable dynamic IR Drop estimation via maximum convolutional neural network [C]. 2020 25th Asia and South Pacific Design Automation Conference(ASP-DAC),13-18.
- [9] CHHABRIA V A, AHUJA V, PRABHU A, et al. Thermal and IR Drop analysis using convolutional encoder-decoder networks [C]. 2021 26th Asia and South Pacific Design Automation Conference ( ASP-DAC ),690-696.
- [10] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9.
- [11] BADRINARAYANAN V, KENDALL A, CIPOLL R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(2015): 2481-2495.
- [12] CHEN L, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrousconvolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, 834-848.
- [13] CHEN S, DE B M. An End-to-end approach to semantic segmentation with 3D CNN and posterior-CRF in medical images [J]. ArXiv Preprint, 2018, ArXiv:1811:03549.
- [14] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [J]. ArXiv Preprint, 2017, ArXiv: 1706:05587.
- [15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. Computer Science, 2014, DOI: 10.48550/arXiv.1409.1556.
- [16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2015): 770-778.
- [17] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops ( CVPRW ),

- IEEE, 2020.
- [18] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutionalblock attention module[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 3-19.
- [19] YANG J, CAI Y, ZHOU Q, et al. Friendly fast poisson solver preconditioning technique for power grid analysis[J]. IEEE Transactions on Very Large Scale Integration Systems, 2014, 22(4):899-912.
- [20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE 2018, 7132-7141.
- [21] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE 2020, 11531-11539.
- [22] PARK J, WOO S, LEE J Y, et al. BAM: Bottleneck attention module[J]. ArXiv Preprint, 2018, ArXiv: 1807:06514.
- [23] CHAI Z, ZHAO Y, LIN Y, et al. CircuitNet: An opensource dataset for machine learning applications in electronic design automation(EDA)[J]. 中国科学:信息科学(英文版), 2022, 65(12): 2. DOI: 10.1007/s11432-022-3571-8.
- [24] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on PatternAnalysis and Machine Intelligence, vol. 39, no. 4, 640-651.
- [25] HASSANIEN M A, SINGH V K, PUIG D, et al. Predicting breast tumor malignancy using deep ConvNeXt radiomics and quality-based score pooling in ultrasound sequences [J]. Diagnostics, 2022, 12(5): 1053.
- [26] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth  $16 \times 16$  words: Transformers for image recognition at scale[J]. ArXiv Preprint, 2020, ArXiv:2010:11929.

### 作者简介

李岳,本科,主要研究方向为深度学习,VLSI 的 EDA 设计。

E-mail:3144596102@qq.com

夏益民(通信作者),博士、副教授,主要研究方向为 AI 算法与集成电路 EDA 设计、移动机器人导航芯片与算法。

E-mail:xiayimin@gdut.edu.cn