基于深度学习的端到端掩模优化任务

汤府鑫[1],徐 辉[2]
1. 安徽理工大学 人工智能学院,安徽 淮南 232001
2. 安徽理工大学 计算机科学与工程学院,安徽 淮南 232001

摘 要:目的 针对光刻系统与特征尺寸不匹配导致的光刻图案与掩模图案严重偏差的问题,提出了一个基于深度学习的端到端掩模优化框架 TransU-ILT。**方法** 该框架使用 CNN-Transformer 的混合模型作为特征提取模块提取目标布局的深度特征,在特征重构模块中加入像素重组层来重构掩模;此外,在训练过程中,加入深度监督机制提高对布局图案特征的提取精度,从而进一步提高掩模的可印刷性。**结果** 实验定量结果表明:与最先进的方法相比,所提出的框架可以实现 4 倍的周转时间加速,在平方 L_2 误差和工艺变化带指标方面分别降低了 13.4%和 4.3%,且框架生成的掩模晶圆图案边缘更加平滑,更接近目标布局。**结论** TransU-ILT 在时间性能和掩模可印刷性方面总体上优于对比的先进方法,可以为掩模优化方法提供一种有效的解决方案。

关键词:掩模优化;光学邻近校正;深度学习;计算光刻
中图分类号:TN305.7 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2024.0006.005

End-to-end Mask Optimization Task Based on Deep Learning

TANG Fuxin[1], XU Hui[2]
1. School of Artificial Intelligence, Anhui University of Science and Technology, Anhui Huainan 232001, China
2. School of Computer Science and Engineering, Anhui University of Science and Technology, Anhui Huainan 232001, China

Abstract: Objective Aiming at the problem of serious deviation between lithography pattern and mask pattern caused by the mismatch between lithography system and feature size, an end-to-end mask optimization framework TransU-ILT based on deep learning was proposed. **Methods** The framework used the CNN-Transformer hybrid model as a feature extraction module to extract the depth features of the target layout, and added a pixel reorganization layer to the feature reconstruction module to reconstruct the mask. In addition, in the training process, the depth supervision mechanism was added to improve the extraction accuracy of layout pattern features, so as to further improve the printability of the mask. **Results** Quantitative experimental results showed that compared with the most advanced methods, the proposed framework achieved a 4X turnaround time acceleration, reduced the square L_2 error and the process change band index by 13.4% and 4.3%, respectively, and the wafer pattern edge of the mask generated by the framework was smoother and closer to the target layout. **Conclusion** TransU-ILT is superior to other advanced methods in terms of time performance and mask printability, which can provide an effective solution for mask optimization methods.

Keywords: mask optimization; optical proximity correction; deep learning; computational lithography

收稿日期:2023-07-17 **修回日期:**2023-09-19 **文章编号:**1672-058X(2024)06-0039-10
基金项目:国家自然科学基金项目资助(62027815,62274052,61404001).
作者简介:汤府鑫(1999—),男,硕士研究生,从事深度学习在 EDA 中的应用研究.
通讯作者:徐辉(1979—),男,博士,教授,从事可信计算与人工智能研究. Email: austxuhui@163.com.
引用格式:汤府鑫,徐辉. 基于深度学习的端到端掩模优化任务[J]. 重庆工商大学学报(自然科学版),2024,41(6):39-48.
TANG Fuxin, XU Hui. End-to-end mask optimization task based on deep learning[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(6): 39-48.

1 引 言

随着半导体制造技术的不断进步,现代电路设计和工艺步骤的复杂性不断提高,掩模优化技术也面临着巨大的挑战。在摩尔定律[1]的推动下,光刻技术跨越了由 90 nm 到 7 nm 及以下的多个工艺节点,逐步逼近分辨率的物理极限,即集成电路的关键尺寸远小于曝光波长[2]。为了补偿由光刻系统限制引起的光刻误差,分辨率增强技术[3]被引入,例如亚分辨率辅助特征[4]插入和光学邻近矫正(Optical Proximity Correction, OPC)[5]技术,可以帮助设计人员获得优化的掩模和高质量的晶圆图像。

光学邻近矫正是一种广泛使用的分辨率增强技术,其基本原理是利用衍射光束之间的相互干涉效应,通过在掩模板上对需要成像的图形形状进行微调,以消除最终在晶圆上成型的图像转角等边缘处可能出现的缺陷,使硅片上的图像更接近目标布局图像。主流 OPC 可以分为 3 类:基于规则的 OPC[6-7]、基于模型的 OPC[8-9]和基于逆光刻技术(Inverse Lithography Technology, ILT)的 OPC[10-14]。

基于规则的 OPC 通过在掩模板上添加规则的辅助图形来修正图像的形变。该技术的优点是计算量相对较小,适用于制造常见的结构和器件。然而该方法非常依赖设计人员的经验,修正规则是从大量实验数据中归纳出来的,并且是在特定照明条件下产生的。如果工艺条件发生改变,则这些规则就需要重新修订。这对于简单模式下的布局图案是快速有效的,但随着光学畸变的加剧,这些规则变得极为复杂而难以继续进行。基于模型的 OPC 从 90 nm 技术节点开始被广泛应用,它利用数学模型来模拟光学和化学反应的影响,然后通过优化算法来调整掩模上的图案,以使其尽可能地接近目标图案。Su 等[9]提出了一种自适应权值的边缘放置误差(EPE)优化模型以调整掩模优化流程中的 EPE 和工艺变化带平衡,并在迭代过程中采用动态分割方法对 EPE 灵敏度高的片段进一步分割为更短的片段,以实现细粒度的 EPE 最小化,从而加快收敛速度,但掩模的复杂性较高,不利于制造。由于目标布局的复杂性日益增加,基于模型的 OPC 方法受到解空间有限和运行时间较长的限制而无法取得令人满意的结果。基于 ILT 的 OPC 采用数值优化思想,将 OPC 任务视为执行像素级掩模优化的逆成像问题;Gao 等[10]结合传统基于模型的掩模优化方法和基于规则的掩模优化方法,同时考虑工艺窗口的影响,并采用计算优化方法减少掩模优化过程中的计算量,从而提高了优化效率,但模型的时间复杂度过高。

随着计算硬件和深度学习技术的蓬勃发展,研究人员开始尝试将 ILT 与 GPU 加速和深度学习模型相结合。Yang 等[11]首次应用条件生成对抗网络(CGAN)来模拟 OPC 过程,它可以从改进的体系结构和目标中学习目标-掩模映射,在收敛时,生成网络能够为给定的目标电路图案创建准最优掩模,并且只需较少的 OPC 步骤来生成高质量的掩模,但生成的掩模对工艺条件的变化很敏感;Jiang 等[12]提出一个在单个神经网络中基于 ILT 的端到端高性能 OPC 框架,以基于学习的方法取代传统的端到端 ILT 过程,该框架由基于预训练模型的掩模预测模块和在线 ILT 校正模块组成,获得了更低的掩模复杂性和更高的掩模可印刷性,但掩模优化的周转时间仍较长;Chen 等[13]将基于水平集的方法迁移到深度神经网络中,建立端到端的联合掩模优化流程,对传统基于水平集的 ILT 算法进行了改进,引入曲率项来降低掩模复杂度,并应用 GPU 加速来克服计算瓶颈,大大减少了运行时间和掩模复杂度,但掩模的可印刷性较差;Wang 等[14]提出一个具有空间注意力机制的 GPU 加速 ILT,通过引入空间注意力图和实时掩模矫正线性化,显著提高了 ILT 质量,并通过强化学习部署增强了该方法的鲁棒性,但优化过程较为烦琐,时间复杂度大。上述深度学习开创性工作的结果令人鼓舞,但它们揭示了一个关键事实,即在大多数与光刻相关的应用场景中,深度学习被视为在结果质量和周转时间之间进行权衡的折中解决方案,计算成本和掩模质量仍不能令人满意。

为了进一步缩短框架训练时间和提高掩模可印刷性,提出一种基于改进 TransUNet 算法[15]的端到端掩模优化框架 TransUNet-ILT。该设计集成了 TransUNet 和 Pixel Shuffle 结构[16],并通过引入深度监督机制[17],实现了从不同深度提取掩模图像内部的语义特征,增强了中间特征图的品质,同时缓解了随着网络规模增加带来的梯度消失问题,从而保证了更好的模型收敛性和掩模可印刷性。

主要贡献:提出了一个基于 TransUnet 的端到端 IL 框架 TransU-ILT,利用 Transformer 模型的全局上下文信息捕捉能力,直接为给定布局生成优化掩模,而不需要进行额外的细化;在训练阶段引入深度监督机制,以提高对布局图案特征的提取精度,从而进一步优化最终的掩模。实验定量结果表明:与 SOTA 方法相比,所提出的框架可以实现 4 倍的周转时间加速,在平方 L_2 误差和工艺变化带指标方面分别降低了 13.4%和 4.3%。

2　准备工作

2.1　光刻模拟

光刻模拟(图 1)是在不进行实际光刻的情况下,模拟输入掩模 D_m 到输出晶圆图像 D_w 的映射。部分相干成像系统的霍普金斯衍射模型被广泛用于近似打印行为[18]。光刻模拟产生一个空间像 D_a。晶圆图像 D_w 是根据晶圆平面上的光强分布生成的,该光强分布由霍普金斯衍射理论的仿真模型产生。通过将掩模 M 与一组预先计算的相干卷积核 h_k 进行卷积并与相应系数 w 进行加权求和获得空间像,如式(1)所示:

$$D_a = \sum_{k=1}^{N^2} w_k \| D_m \otimes h_k \|_2^2 \tag{1}$$

其中,D_m 表示掩模图案,h_k 和 w_k 是光刻仿真模型的第 k 个内核和相应权重,$\otimes$表示卷积运算。通过截断求和获得相干成像系统的降阶近似[10]。则式(1)可以表示为

$$D_a = \sum_{k=1}^{N_h} w_k \| D_m \otimes h_k \|_2^2 \tag{2}$$

式(2)中,$N_h = 24$。得到光强分布之后,空间像经过抗蚀剂模型用来评估晶圆上的最终印刷图案。抗蚀剂模型由以下阈值函数表示:

$$D_w = \begin{cases} 1, \text{ if } D_a(x,y) \geqslant th_a \\ 0, \text{ if } D_a(x,y) < th_a \end{cases} \tag{3}$$

在实验中,式(3)设置 $th_a = 0.225$。通过上面函数可以得到二值的晶圆图像,但在处理二值晶圆图像时,由于像素值是固定的,因此计算会出现损失,导致梯度损失,从而不利于训练框架参数的更新。为了解决这个问题,采用 sigmoid 函数代替传统的硬阈值函数。sigmoid 函数将光刻强度图转换为每个像素的数字在 0~1 之间变化的晶圆图像,而不是仅仅是 0 或 1。

$$D_w(x,y) = Sig(D_a(x,y)) = \frac{1}{1+e^{-\theta(D_a(x,y)-th_a)}} \tag{4}$$

式(4)中,θ 表示 sigmoid 函数的陡峭程度。如文献[10]中所建议的,设置 $\theta = 50$,$th_a = 0.225$。这种方式使得在训练框架参数时,梯度能够更好地传播,从而提高了训练的效率和准确性。

图 1　光刻模拟过程

Fig. 1　Lithography simulation process

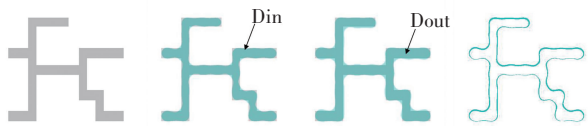
2.2　评价方法

掩模优化的最终目标是改善掩模的可印刷性,使得光刻过程中晶圆图案与布局图案的偏差尽可能小,以达到高质量、高精度的制造目标。为了衡量掩模和布局图案之间的差异,可以使用平方 L_2 误差和工艺变化带进行度量,其定义如下:

定义 1(平方 L_2 误差):平方 L_2 误差由 $\| D_w - D_t \|_2^2$ 计算,其中 D_t 为目标布局图案,D_w 为标称工艺条件下生成的晶圆图案。

在工艺制造中,焦点深度和剂量通常被视为工艺变量,它们可以影响光刻过程的结果。可以通过工艺变化带来帮助评估工艺变化对光刻结果的影响,从而优化光刻参数并提高制造质量。

定义 2(工艺变化带):工艺变化带(PV Band)是通过最小和最大工艺条件下两个生成的晶圆图像 D_{in} 和 D_{out} 之间按位异或求和得出的。如图 2 所示,焦点和剂量的变化会引起图像失真和轮廓偏差。PV Band 区域越小,掩模对工艺变化的敏感度越低,容忍度越高。



(a) 目标布局　(b) 散集,低剂量　(c) 散集,高剂量　(d) 工艺变化带

图 2　工艺变化带示例图

Fig. 2　Example diagram of a process variation band

根据以上描述,掩模优化任务定义如下:

问题 1(掩模优化)　给定目标布局 D_t,使得从基于模型 $f(\cdot;P_{nom})$ 模拟的晶圆图像具有较小的平方 L_2 误差、PV Band 和掩模优化周转时间,从而实现高质量、高效率的制造目标。

3　TransU-ILT 算法

将整个 ILT 过程视为一个黑盒,学习输入目标布局和输出优化掩模之间的映射可以自然地表述为图像

到图像的转换任务。在语义分割任务中,模型需要将输入图像转换成与其大小相同的分割图像,这也可以看作是一个图像到图像的转换过程。传统的语义分割任务基于 CNN 来提取图像特征,但是卷积操作具有局部性质,难以捕获全局特征,导致在处理一些全局信息比较重要的任务时表现不佳。为了解决这一问题,可以使用多头注意力机制来增强网络的特征提取能力。受到文献[15]的启发,将 Transformer 结构引入到掩模优化任务中,设计了一种端到端的掩模优化框架 TransU-ILT。

3.1 总体架构

TransU-ILT 的总体架构如图 3 所示。如图 3(b)左侧所示,该特征提取模块基于 TransUnet[15] 设计,它不是使用纯 Transformer[19] 作为编码器,而是使用 CNN-Transformer 的混合模型。图 3(b)的右侧是特征重构模块,包括深度监督机制和像素重组层。

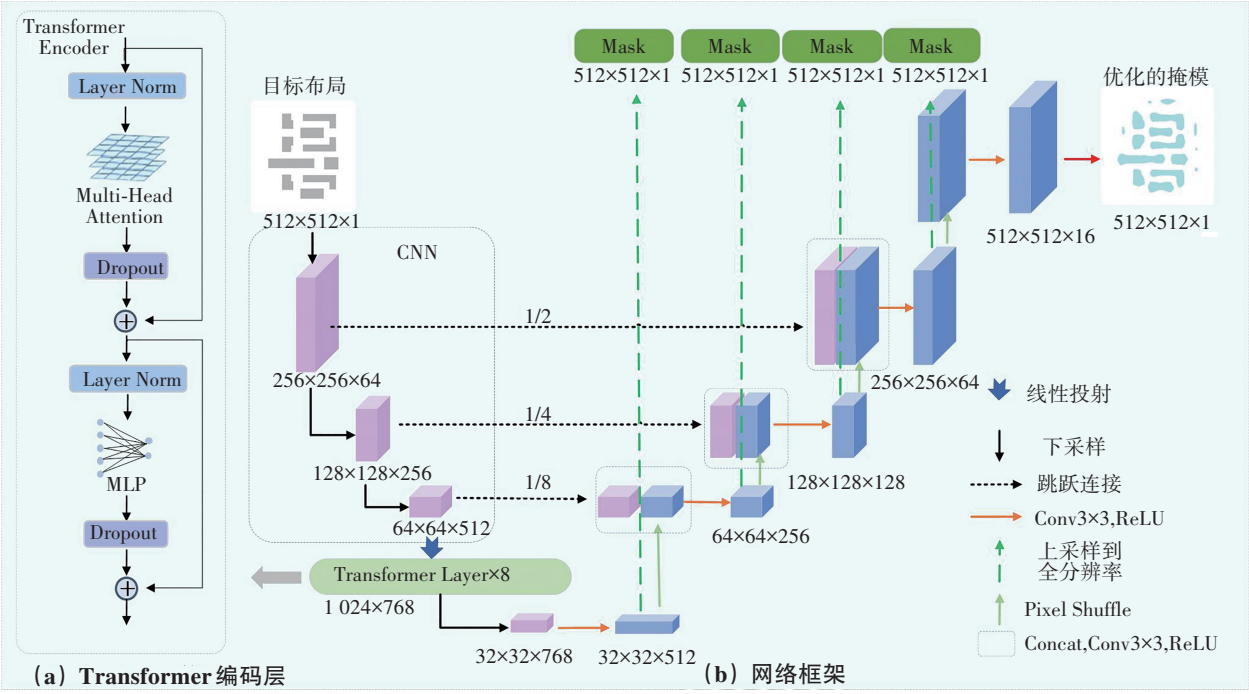


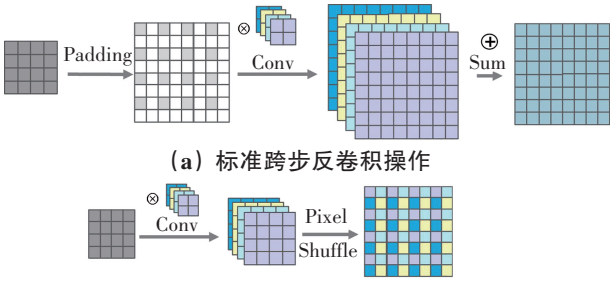
图 3 整体框架

Fig. 3 Overview framework

对于给定目标布局 $D_t \in R^{H\times W\times C_{in}}$($H$、$W$、$C_{in}$ 分别表示布局图像的高度、宽度和输入通道数。H 和 W 为 512,由于布局图像是一个二值图像,所以 $C_{in}=1$。),首先使用 CNN 提取浅层特征,D_t 经过一层卷积得到 256×256×64 大小的特征图,然后使用 ResNet50 模型[20] 提取浅层特征得到 32×32×1024 大小的特征图。浅层特征包含边缘信息以及目标布局和对应掩模的细节,是掩模优化的关键特征。接下来是 Transformer 主体,包括 8 个 Transformer 模块,用于从浅层特征中提取深层特征,利用 Transformer 模型强大的全局上下文信息捕捉能力提取布局图案的深度特征,经过 Transformer 层后深度特征变为 $F_d \in R^{32\times32\times768}$,然后将其送入到卷积层,将提取的特征进行非线性映射为 $F_{df} \in R^{32\times32\times512}$,进而提高模型的泛化能力。

接下来是特征重构模块,该特征重构模块将特征 $F_{df} \in R^{32\times32\times512}$ 重构为优化掩模 $D_m \in R^{512\times512\times1}$。由于 32×32 比原始图像分辨率 512×512 小得多,因此不可避免地会导致低级特征信息(如目标布局的边缘、顶点和拐角)的丢失。为了弥补这种信息损失,如图 3 所示,采用跳跃连接的策略,在编码器和解码器之间建立连接,将编码器中低级别特征图与解码器中高级别的特征图进行连接。这种连接方式能够有效地结合低级别特征图的空间上下文信息和高级别特征图的语义信息,从而更好地还原图像的细节信息。特征重构最主要的就是对低分辨率特征图进行上采样,在传统方法中,通过如图 4(a)所示的标准跨步反卷积操作,在现有的像素之间插入 0,使得在卷积运算之后的输出分辨率达到期望值。然而,这种机制需要在高分辨率空间上进行多次卷积运算,计算效率不高,而且可能会产生棋盘状伪影等问题,如图 5(a)所示。而 Pixel Shuffle[16] 是另一种在超分辨率任务中广泛使用的上采样解决方案,如图 4(b)所示,它将形状为 $H\times W\times r^2C$ 的张量投射成形状 $rh\times rW\times C$。它比反卷积运算更快地提高了特征映射的

分辨率,最终产生高分辨率的掩模图像。同时,Pixel Shuffle 只对特征像素进行重新排序,而不使用卷积核和池化,避免了噪声的增加,从而使得每个像素都具有较高的信息量,如图 5(b)所示。经过 4 层 Pixel Shuffle 块和 3×3 卷积操作,得到最终的优化掩模 $D_m \in R^{512\times512\times1}$。

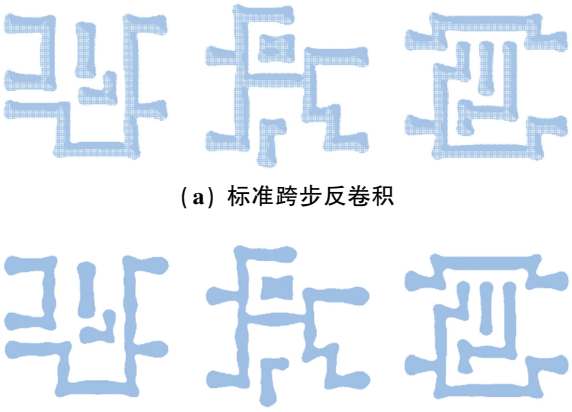


(a) 标准跨步反卷积操作

(b) Pixel Shuffle 反卷积操作

图 4　反卷积操作

Fig. 4　Deconvolution operation



(a) 标准跨步反卷积

(b) Pixel Shuffle 反卷积

图 5　反卷积操作生成的图像

Fig. 5　Pattern generated by deconvolution operation

3.2　Transformer Block

Transformer 模块[21]最初是在自然语言处理中使用的,用于处理序列数据,为了使 Transformer 也适用于掩模优化任务,需要将 Transformer 模块的输入从序列数据改为图像数据。使用多头自注意力机制来处理输入序列中的不同特征,并分别对它们进行编码。具体来说,将小尺寸特征图划分为图像块 $I_1, I_2, \cdots, I_i \in R^{P^2\times C}$, $i=1,2,\cdots,N$,每个图像块的尺寸均为 $P\times P$,图像块总数 $N=H\times W/P^2$(H、W 为图像的高度和宽度)。将 I_i 进行如式(5)所示的线性投射,

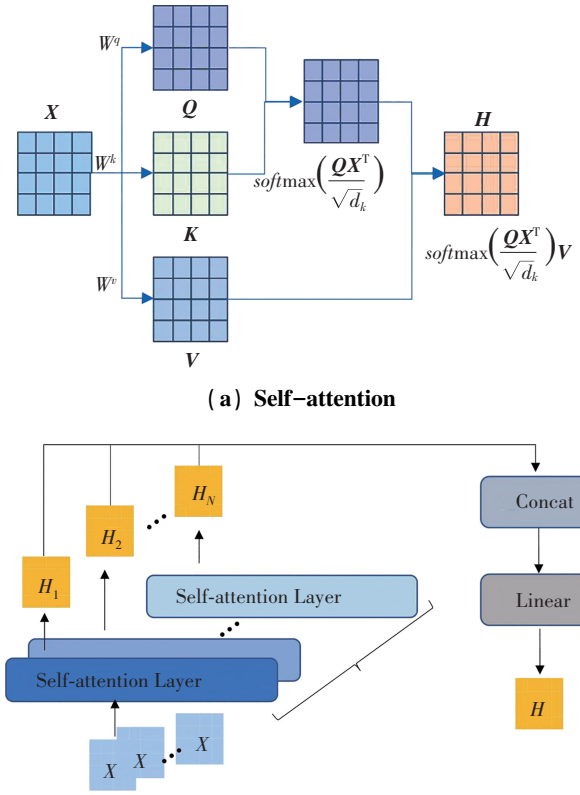
$$\boldsymbol{F}_0=[I_1\boldsymbol{E}, I_2\boldsymbol{E}, \cdots, I_N\boldsymbol{E}]+E_{pos} \tag{5}$$

其中,E 为线性映射矩阵,E_{pos} 为位置编码,F_0 是线性映射后的向量,是 Transformer 模块的输入。如图 3(a)所示,Transformer 模块的核心部件由 MSA 多头自注意力机制和 MLP 前馈神经网络组成。具体流程:首先是 Layer Norm 层,对输入进行标准化,使得每个特征维度的均值为 0,方差为 1,防止输入的变化过大,导致网络训练不稳定;接下来是多头自注意力机制(如图 6 所示),用于捕捉输入的全局信息,在这里,它可以同时处理多个位置的信息,并得到不同的注意力权重,从而增强特征提取能力;然后对多头自注意力机制的输出进行 Dropout 随机失活后,将其输出和输入相加,防止梯度消失并保留输入的信息;然后是 MLP 前馈神经网络,对多头自注意力机制的输出进行非线性变换,增强模型的表达能力,与多头自注意力机制类似,对前馈神经网络的输出也进行残差连接和层归一化,以提高模型的稳定性和泛化能力。所以,第 l 层的输出如下所示:

$$F_l'=MSA(LN(F_{l-1}))+F_{l-1}$$

$$F_l=MLP(LN(F_l'))+F_l'$$

其中,$MSA(\cdot)$表示多头自注意力模块,$LN(\cdot)$表示层归一化算子。



(a) Self-attention

(b) Multi-headed Self-attention

图 6　多头自注意力机制

Fig. 6　Multi-headed self-attention mechanism

3.3　深度监督机制

深度监督机制通过在深层隐藏层中添加额外的目标函数来监督特征图的质量。Lee 等[17]提出了一种名为 DSN (Deeply-Supervised Nets)的结构,通过在深度神经网络的隐藏层中加入监督信息来评估隐藏层中产

生的特征图质量,在图像分类任务中取得了较好的效果。对于掩模优化任务来说,不同的布局图案边缘、拐角和布局之间的距离信息是不同的,如果在浅层的隐藏层中添加额外的深度反馈,可以进一步优化深层输出特征图的质量,进而优化最终的掩模。另一方面,仅靠输出层在图像原分辨率下进行监督会使网络侧重细节纹理信息的学习,而对于较深的隐藏层进行监督,可以直接让网络学习目标布局的语义信息,从而提高模型不同布局图案下的优化效果。如图 3 所示,将每一个深度的特征张量都上采样到输入分辨率大小(512×512)进行深度监督,相应的目标函数为

$$L_{\text{deep}} = \| D_1 - D_t \|_2^2 + \| D_2 - D_t \|_2^2 + \| D_3 - D_t \|_2^2 + \| D_4 - D_t \|_2^2 \tag{6}$$

式(6)中,D_1、D_2、D_3、D_4 分别为在不同深度下掩模的晶圆图案,D_t 为目标布局。

3.4 损失函数

基于学习的方法中,损失函数是训练目标。在传统方法中,训练框架通常最大限度地降低以下目标:

$$L = \alpha \cdot \| D_w - D_t \|_2^2 + \beta \cdot \| f(D_m; R_{\max}) - f(D_m; R_{\min}) \|_2^2 \tag{7}$$

其中,D_w 是框架生成的掩模对应的晶圆图像,D_t 是数据集中相应的目标布局,$f(D_m; R_{\max})$ 和 $f(D_m; R_{\min})$ 为最大和最小工艺条件下的晶圆图像。为了保证优化算法对不同工艺条件的鲁棒性,需要考虑工艺变化感知的损失函数。受文献[22]启发,将工艺变化的损失函数变为

$$L_{\text{pvb}} = \| f(D_m; R_{\max}) - D_t \|_2^2 + \| f(D_m; R_{\min}) - D_t \|_2^2$$

为了使式(7)中的损失值更低,要求 $f(D_m; R_{\max})$ 和 $f(D_m; R_{\min})$ 都接近目标布局,而不仅仅是最小化它们之间的距离,这种要求对优化过程施加了更加严格的限制。调整后的损失函数可以指导优化过程,找到所需的掩模,从而生成 PV Band 可接受的晶圆图像。如 3.3 节所述,在模型中加入深度监督机制,同时为了加快模型的收敛速度,将目标布局与其预先标记的掩模间的损失添加到损失函数中,调整后框架的目标函数为

$$L = \alpha \| D_w - D_t \|_2^2 + \beta (\| f(D_m; R_{\max}) - D_t \|_2^2 + \| f(D_m; R_{\min}) - D_t \|_2^2) + \gamma \| D_m - M^* \|_2^2 + \mu L_{\text{deep}} \tag{8}$$

其中,α、β、γ、μ 是可配置的超参数。由于目标布局对应的预先标记的掩模不是严格光刻下的掩模,因此,γ 只设置了一个较低的值。

4 仿真实验与结果分析

4.1 实验细节

TransU-ILT 框架采用 Pytorch 和 CUDA 开发,并在一台配备有 2.50 GHz Inter Xeon CPU 和单个 NVIDIA RTX3090 GPU 的 Linux 机器上进行测试。光刻模拟模型是从 Neural-ILT[12] 中获得的基于 CUDA 的高性能光刻仿真模型。训练集是从 GAN-OPC[11] 中获得的基于 32 nm M1 布局的训练数据集,有 4 875 对。评估数据集是由 ICCAD 2013 竞赛[23]发布的不同形状的矩形和多边形组成的十个基准,每个基准布局是一个 32 nm 的 M1 金属层。图像尺寸为 2 048 nm×2 048 nm,每像素分辨率为 1 nm^2。采用 193 nm 波长的光刻系统,离焦范围为±25 nm,剂量范围为±2%,这是由 ICCAD 2013 竞赛[23]提供的。最外面的最终印刷图案在标称聚焦和+2%剂量下产生,最里面的印刷图案在散焦和-2%剂量下产生。光刻胶模型(式 4)的阈值强度 $th_r = 0.225$。模型输入分辨率为 512×512,用 Adam 优化器对模型进行训练,其学习率为 0.000 2,动量为 0.9,权重衰减为 0.000 1。训练目标式(8)的超参数选择为 $\alpha = 2$,$\beta = 1$,$\gamma = 0.1$,$\mu = 0.2$。TransU-ILT 经过 10 个 epoch 的训练,在单个 RTX3090 GPU 上大约需要 3.5 h。

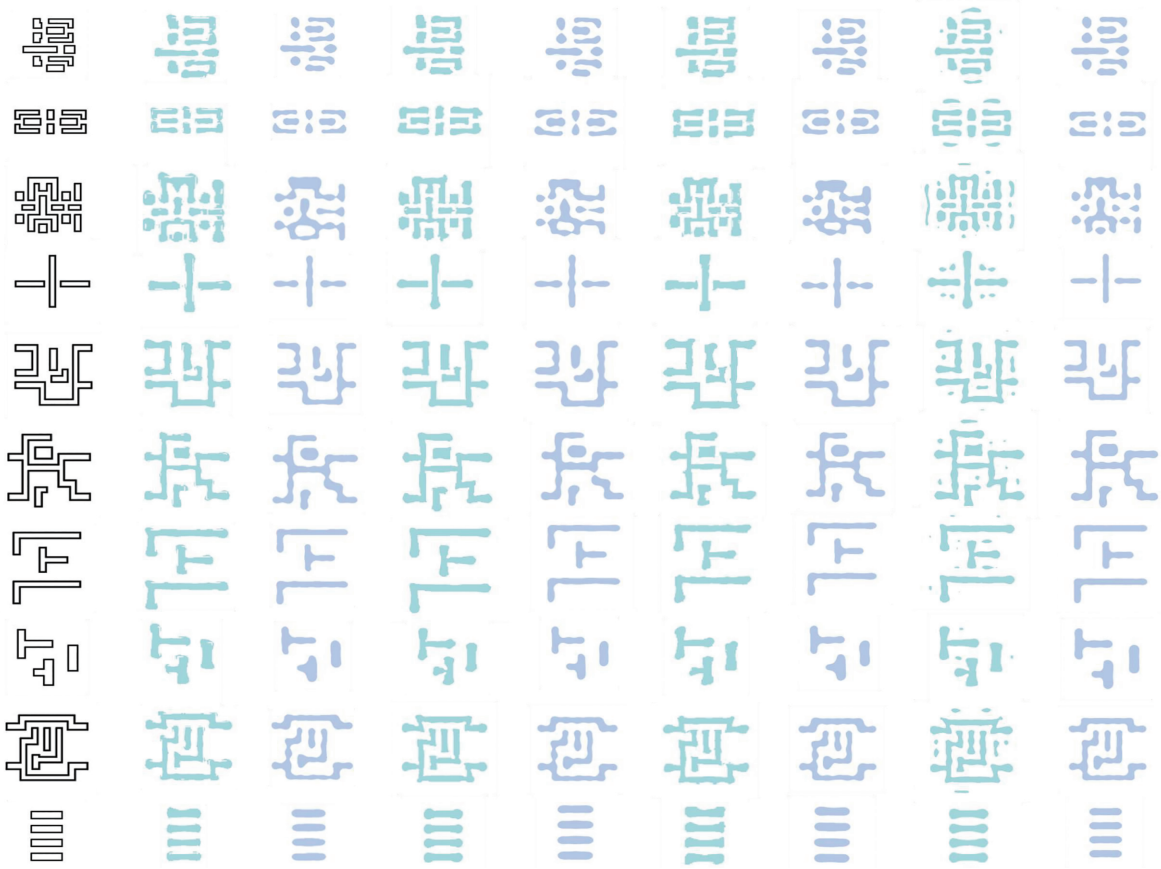
4.2 TransU-ILT 的评估

为了验证所提出 TransU-ILT 的有效性,对 ICCAD 2013 竞赛提供的 10 个工业布局进行优化测试,其分辨率为 2 048×2 048,由于计算的复杂性,将其缩减为 512×512 输入到框架中。与传统 ILT 方法[10]和基于学习的 SOTA 方法(PGAN-OPC[11]、Neural-ILT[12]和 A2-ILT[14])相比,定量结果如表 1 所示。值得注意的是 A2-ILT 矫正结果从 https://github.com/cuhk-eda/A2-ILT 获得,其中"时间"列表示在没有训练时间的情况下,从给定目标布局到优化掩模的周转时间;"L_2"(单位:nm^2)列显示优化掩模的晶圆图像与目标布局之间的平方 L_2 误差;"PVB"(单位:nm^2)列表示在±2% 剂量误差和聚焦或散焦下的工艺变化带。可以观察到,TransU-ILT 在"TAT"、"L_2"和"PVB""指标方面都优于其他方法。更具体地说,与 ILT、PGAN-OPC、Neural-ILT 和 A2-ILT 相比,TransU-ILT 分别实现了 596 倍、281 倍、9 倍和 4 倍的 TAT 加速,减少了 35.4%、22.9%、15.4%和 13.4%的平方 L_2 误差,减少了 4.3%、2.4%、4.4%和 4.3%的工艺变化带。SOTA 方法和本文提出框架生成的 10 个掩模及其晶圆图像如图 7 所示。与其他方法相比,该框架生成的掩模晶圆图案边缘更加平滑,且更接近目标布局。此外,可以观察到掩模边缘存在一些细小的斑点,这些斑点是模型自动添加的辅助特征,有助于提高掩模的可印刷性。

表 1　本文方法与现有方法的性能比较结果

Table 1　Performance comparison results of the proposed method and the state-of-the-art methods

基　准		ILT[10]			PGAN-OPC[11]			Neural-ILT[12]			A2-ILT[14]			TransU-ILT		
标号	面积/nm^2	时间/s	L_2	PVB	时间/s	L_2	PVB	时间/s	L_2	PVB	时间/s	L_2	PVB	时间/s	L_2	PVB
1	215 344	1 280	49 893	65 534	358	52 570	56 267	11	49 817	55 975	6.18	45 287	59 940	1.76	44 132	52 551
2	169 280	381	50 369	48 230	368	42 253	50 822	17	38 174	52 010	5.59	34 044	51 988	1.41	32 097	43 162
3	213 504	1 123	81 007	108 608	368	83 663	94 498	10	89 411	91 357	5.67	92 505	91 261	1.41	85 323	88 145
4	82 560	1 271	20 044	28 285	377	19 965	28 957	9	16 744	29 982	5.63	21 644	29 017	1.23	13 582	29 652
5	281 958	1 120	44 656	58 835	369	44 733	59 328	11	45 598	58 900	5.68	38 082	61 601	1.25	31 733	59 009
6	286 234	391	57 375	48 739	364	46 062	52 845	10	43 836	54 969	5.64	42 068	53 620	1.23	35 537	54 372
7	229 149	406	37 221	43 490	377	26 438	47 981	16	20 324	50 542	5.64	21 947	49 053	1.25	21 204	50 590
8	128 544	388	19 782	22 846	383	17 690	23 564	15	13 337	26 353	5.52	15 668	23 853	1.22	14 417	25 985
9	317 581	1 138	55 399	66 331	383	56 125	65 417	11	49 401	68 817	5.59	46 973	68 442	1.25	38 592	64 952
10	102 400	387	24 381	18 097	366	9 990	19 893	14	8 511	20 734	5.44	10 450	19 950	1.22	8 430	19 562
平均值	—	788.5	44 012.7	50 899.5	371.3	39 948.9	49 957.2	12.4	37 515.3	50 963.9	5.658	36 866.8	50 872.5	**1.323**	**32 504.7**	**48 798**
比率	—	595.994	1.354	1.043	280.650	1.229	1.024	9.373	1.154	1.044	4.277	1.134	1.043	**1.000**	**1.000**	**1.000**



(a) 目标布局　(b) PGAN-OPC　(c) Neural-ILT　(d) A2-ILT　(e) TransU-ILT

图 7　本文模型与对比模型的掩模及其晶圆图案效果对比图

Fig. 7　Comparison of masks and their wafer pattern effects between the proposed model and the comparison model

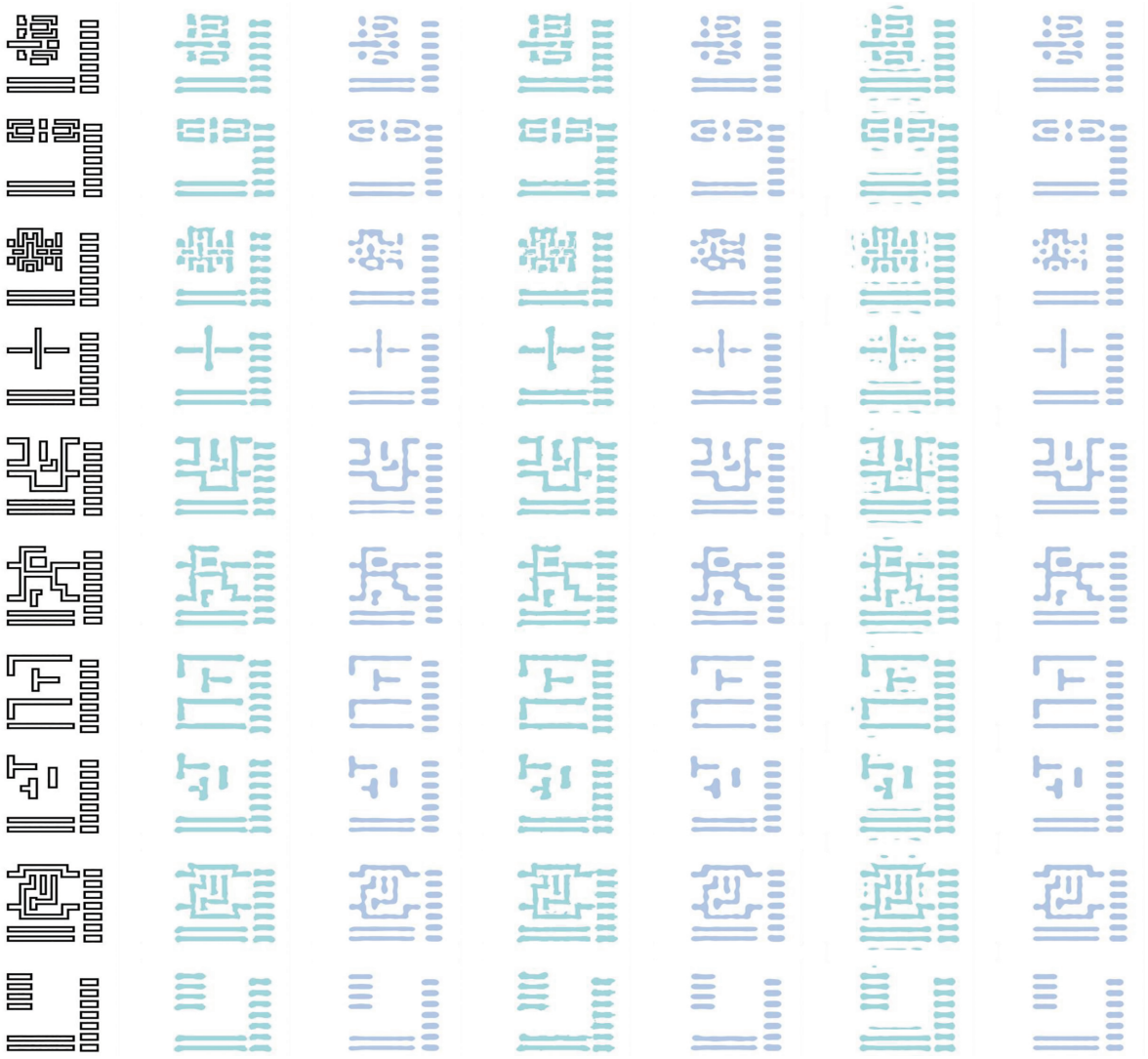
4.3　TransU-ILT 的稳定性

为了进一步验证框架的稳定性,在另外 10 个包含更多图案的测试用例上进行了进一步的实验。这 10 个测试用例都是由具有附加几何图形的原始 IBM 基准创建的。定量结果如表 2 所示。可以观察到,与 Neural-ILT[12] 和 A2-ILT[14] 相比,TransU-ILT 分别实现了 13 倍和 4 倍的周转时间加速,减少了 19.2% 和 20.6% 的平方 L_2 误差,减少了 6.7% 和 8.5% 的工艺变化带。本文提出的框架相对于其他方法,在大型基准测试中获得了更好的性能。用最先进的方法和 TransU-ILT 生成的 10 个大型掩模及其晶圆图像效果对比如图 8 所示。相较于其他方法,该框架生成的掩模晶圆图案边缘更加平滑,且更接近目标布局。因此,本文提出的框架可以更好地推广到更复杂的布局图案。

表 2 在大型基准上与最先进方法的性能比较

Table 2 Performance comparison with state-of-the-art methods on large benchmarks

基准	Neural-ILT[12]			A2-ILT[14]			TransU-ILT		
标号	t/s	L_2	PVB	t/s	L_2	PVB	t/s	L_2	PVB
11	20	79 933	120 577	6.45	77 929	119 661	1.72	69 618	104 185
12	12	86 995	104 266	5.54	73 156	111 214	1.29	56 033	97 505
13	15	133 281	152 718	5.72	127 969	172 232	1.24	115 404	144 825
14	14	43 797	92 137	5.61	56 168	90 744	1.25	40 192	84 513
15	19	69 521	122 115	5.6	77 822	116 935	1.24	61 276	110 431
16	19	73 790	117 359	5.68	80 503	111 049	1.22	66 547	107 278
17	19	49 031	92 320	5.71	48 239	91 066	1.25	50 073	96 611
18	16	47 409	84 971	5.69	49 331	84 718	1.27	39 045	79 395
19	14	93 922	115 028	5.91	85 551	123 373	1.22	68 236	115 844
20	19	38 028	80 127	5.58	47 690	79 608	1.25	34 238	73 484
平均值	16.7	71 571	108 162	5.7	72 436	110 060	**1.30**	**60 066**	**101 407**
比率	12.846	1.192	1.067	4.385	1.206	1.085	1.00	1.00	1.00



(a) 目标布局 (b) Neural-ILT (c) A2-ILT (d) TransU-ILT

图 8 本文模型与对比模型在大型基准上的掩模及其晶圆图案效果对比图

Fig. 8 Comparison of masks and wafer pattern effects between the proposed model and the comparison model on large benchmarks

4.4　结果分析

4.4.1　跳跃连接的数量

如 3.1 节所述,采用跳跃连接的策略,结合低级别特征图的空间上下文信息和高级别特征图的语义信息,有助于更好地还原图像的细节信息。这个消融的目的是测试在 TransU-ILT 中添加不同数量跳跃连接的影响。通过改变,跳跃连接的数量为 0/1/3,10 个测试目标布局的平均 L_2 误差和工艺变化带性能如图 9 所示。在跳跃连接数量为 1 时,只在 1/4 分辨率处添加跳跃连接。可以看到,添加更多的跳跃连接通常会获得更好的掩模可印刷性。最佳的平均 L_2 误差和 PVB 是在 1/2、1/4 和 1/8 分辨率处加入跳跃连接(如图 3 所示)。因此,TransU-ILT 采用的是这种配置。值得一提的是,复杂布局图案(如 case1 和 case3)的性能增益比简单的布局图案(如 case4 和 case10)更为明显。这些结果证明了加入跳跃连接可以学习到掩模中精确的低级细节。

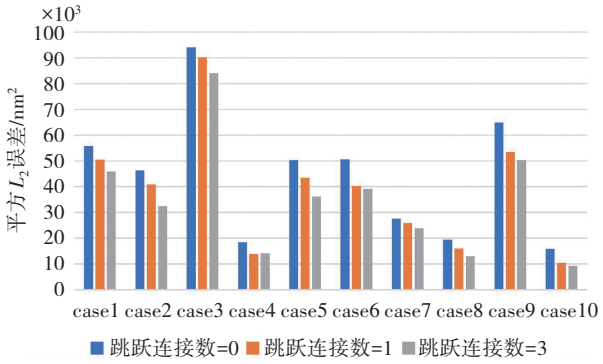


图 9　TransU-ILT 中跳跃连接数量的消融实验

Fig. 9　Ablation experiment of the number of jump connections inTransU-ILT

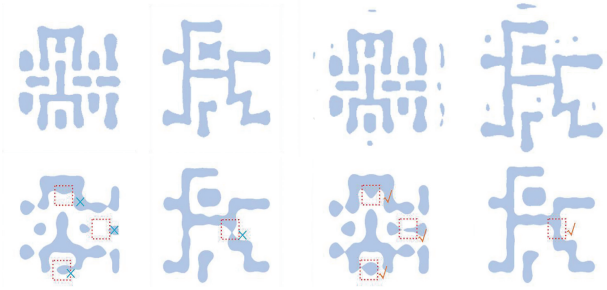
4.4.2　深度监督机制的有效性

如 3.3 节所述,在浅层的隐藏层中添加额外的深度反馈,可以进一步优化深层输出特征图的质量,进而优化最终的掩模。为了验证深度监督机制的有效性,还给出了没有加入深度监督机制的结果,如表 3 所示。可以观察到:加入深度监督机制可以提高掩模的可印刷性。如图 10 所示,分别展示了没有加入深度监督机制和加入深度监督机制的情况下,case3 和 case6 的掩模和晶圆图案。可以观察到:加入深度监督机制的掩模在克服光学临近效应方面表现更好,生成的晶圆图案更贴近目标图案。具体而言,图 10(b)中的晶圆图案更加平滑且形状更加规整,与目标图案更加相似。因此,加入深度监督机制可以有效提升模型的性能和掩模的可印刷性。

表 3　深度监督机制的消融实验

Table 3　Ablation experiments with deep supervision mechanisms

模　型	L_2/nm²	PVB/nm²	L_2+PVB/nm²
TransU-ILT(有深度监督)	**34 826**	**48 394**	**83 220**
TransU-ILT(无深度监督)	35 962	49 569	85 531



(a) 无深度监督　　(b) 有深度监督

图 10　深度监督机制可视化

Fig. 10　Visualization of the deep supervision mechanism

5　结论与展望

在本文中,提出了一种端到端的基于学习的光学邻近校正框架 TransU-ILT,利用 Transformer 模型强大的全局上下文信息捕捉能力,并引入深度监督机制,直接为给定的布局生成优化掩模。评估结果表明:TransU-ILT 在可印刷性方面优于其他 SOTA 解决方案,并且能够显著缩短周转时间,具有一定的实用和推广价值。

未来的工作将探索如何进一步应用图像生成网络来提高掩模优化的性能和效率。此外,还考虑使用更多的数据来训练 TransU-ILT 模型,以提高其泛化能力和鲁棒性。另外,考虑将 TransU-ILT 框架应用于其他类似的问题领域,以发掘其更广泛的应用价值。

参考文献(References):

[1]　MOORE G. Cramming more components onto integrated circuits[J]. Proceedings of the IEEE, 1998, 86(1): 82-85.

[2]　马旭, 张胜恩, 潘毅华, 等. 计算光刻研究及进展[J]. 激光与光电子学进展, 2022, 59(9): 122-170.
MA Xu, ZHANG Shen-en, PAN Yi-hua, et al. Research and progress of computational lithography[J]. Laser & Optoelectronics Progress, 2022, 59(9): 122-170.

[3]　MA X. Optimization of resolution enhancement techniques in

optical lithography[M]. University of Delaware, 2009.

[4] GENG H, YANG H, MA Y, et al. SRAF insertion via supervised dictionary learning[C]//Proceedings of the 24th Asia and South Pacific Design Automation Conference. 2019: 406-411.

[5] OTTO O W, GAROFALO J G, LOW K K, et al. Automated optical proximity correction: a rules-based approach[C]// Proceedings of the SPIE Proceedings, Optical/Laser Microlithography VII, SPIE, 1994.

[6] DONG X, ZHANG L. Process-variation-aware rule-based optical proximity correction for analog layout migration[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2017, 36(8): 1395-1405.

[7] PARK J S, PARK C H, RHIE S U, et al. An efficient rule-based OPC approach using a DRC tool for 0.18 μm ASIC[C]// Proceedings of the 1st International Symposium on Quality of Electronic Design. Piscataway: IEEE Press, 2000: 81-85.

[8] KUANG J, CHOW W K, YOUNG E F Y. A robust approach for process variation aware mask optimization[C]// Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015. New ersey: IEEE Conference Publications, 2015: 1591-1594.

[9] SU Y H, HUANG Y C, TSAI L C, et al. Fast lithographic mask optimization considering process variation[C]// Proceedings of the IEEE/ACM International Conference on Computer-Aided Design. Piscataway: IEEE Press, 2014: 1345-1357.

[10] GAO J R, XU X, YU B, et al. MOSAIC: mask optimizing solution with process window aware inverse correction[C]// Proceedings of the 51st Annual Design Automation Conference. Piscataway: IEEE Press, 2014: 1-6.

[11] YANG H, LI S, MA Y, et al. GAN-OPC: mask optimization with lithography-guided generative adversarial nets[C]// Proceedings of the 2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC). Piscataway: IEEE Press, 2018: 1-6.

[12] JIANG B, LIU L, MA Y, et al. Neural-ILT 2.0: Migrating ILT to domain-specific and multitask-enabled neural network[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2022, 41(8): 2671-2684.

[13] CHEN G, YU Z, LIU H, et al. DevelSet: deep neural level set for instant mask optimization[C]//Proceedings of the IEEE/ACM International Conference on Computer Aided Design. Piscataway: IEEE Press, 2021: 1-9.

[14] WANG Q, JIANG B, WONG M D F, et al. A2-ILT: GPU accelerated ILT with spatial attention mechanism[C]// Proceedings of the 59th ACM/IEEE Design Automation Conference. NewYork: ACM, 2022: 967-972.

[15] CHEN J, LU Y, YU Q, et al. TransUNet: Transformers make strong encoders for medical image segmentation[EB/OL]. [2023-07-27]. https://arxiv.org/abs/2102.04306v1.

[16] SHI W, CABALLERO J, HUSZAR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1874-1883.

[17] LEE C Y, XIE S, GALLAGHER P, et al. Deeply-supervised nets[C]//Artificial intelligence and statistics. PMLR, 2015: 562-570.

[18] HOPKINS H H. The concept of partial coherence in optics[J]. Proceedings of the Royal Society of London Series A Mathematical and Physical Sciences, 1951, 208(1093): 263-277.

[19] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16 × 16 words: Transformers for image recognition at scale[J/OL]. [2023-07-27]. https://arxiv.org/abs/2010.11929, 2020.

[20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.

[21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30: 5998-6008.

[22] YU Z, CHEN G, MA Y, et al. A GPU-enabled level-set method for mask optimization[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2023, 42(2): 594-605.

[23] BANERJEE S, LI Z, NASSIF S R. ICCAD-2013 CAD contest in mask optimization and benchmark suite[C]//Proceedings of the International Conference on Computer-Aided Design. Piscataway: IEEE Press, 2013: 271-274.

责任编辑:李翠薇