基于复杂卷积网络和数据增强的晶圆缺陷模式识别[[1]](#footnote-1)[[2]](#footnote-2)

**摘要**：晶圆缺陷模式识别是IC电路设计优化中的重要辅助手段，在引入自动识别方法之前，工业届依赖于专家进行人工识别，此项工作耗时耗力。近年来，关于晶圆缺陷模式自动识别工作逐渐出现。我们以前人论文为入手点，复现了前人在数据集上的工作，并通过引入复杂卷积神经网络和GAN数据增强的方法，优化了原模型在全测试集上的运行效果，取得了比较不错的结果。

**关键词**： 晶圆图 缺陷模式 模式识别 卷积神经网络 数据增强 GAN

# 引入

## 背景介绍

在IC电路的设计过程中，会遇到一些工艺原因导致的IC缺陷，在实际的排查中，获取晶圆投影得到的晶圆图，并以此作为依据判断晶圆对应的缺陷类型，是辅助排查的重要手段。识别缺陷类型的工作，通常由专业的工程师来完成，对于一批存在IC缺陷的电路，其涉及的晶圆图往往数量巨大，缺陷模式复杂，这项工作对于人工而言过于冗长。近年来，出现了一系列将自动识别引进这一领域的工作，其中比较具有代表性的工作有：2019年S.C.Wong等人完成的将机器学习引入缺陷模式识别，2020年M.B.Alawieh等人使用SelectiveNet的工具在数据集上以90%左右的选择率达到了95%以上的验证集准确率，2022年P.P.Shinde等人运用YOLO大模型对缺陷进行定位，并达到了96%以上的准确率。这些工作都取得了比较好的效果，但对于其中一些模型，均存在着模型较大（YOLO），或者模型的预测覆盖率不高的问题（SelectiveNet），这对于工业届并不是什么大问题，但一个更好的平衡是我们希望能够达到的，如此也可以更加轻量、更加高覆盖率地应用到实际工作中。

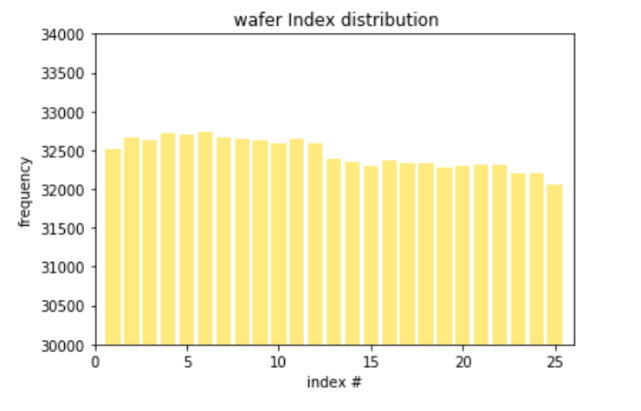
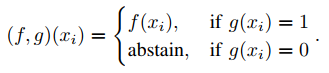
## 研究介绍

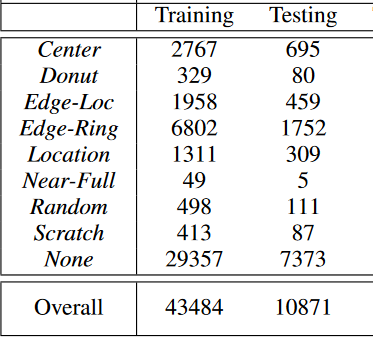
我们的研究对象为晶圆图的模式识别，以M.B.Alawieh等人关于SelectiveNet的工作为入手点，对原始论文进行复刻，并给出改进，进一步引入较为复杂的卷积神经网络，以及通过GAN来达成数据增强的效果。在最后，可以以100%的预测覆盖率，在数据集的测试集上达到96%以上的预测准确率。

# 数据集

我们采用的数据集是晶圆图缺陷模式识别领域常用的WM-811K数据集。

## WM-811K数据集

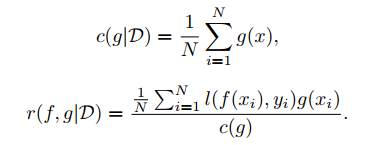
WM-811K数据集发布于2019年，其包含了从46393个槽上收集而来的811467张晶圆图。每一个槽上有25个晶圆，收集到的晶圆图Index分布均匀。所有数据均收集于实际的工业流程中，并由专业工程师进行数据分类。数据集的可信度较高。在使用SelectiveNet进行晶圆缺陷模式识别的论文中，研究者团队使用了这一数据集中的测试集部分，并将其划分为测试集和验证集，这一部分的数据如下图所示

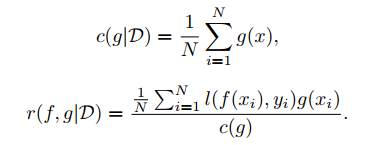


# 初步探索

## Selective Net简介

原论文中，引入SelectiveNet的动机在于对于识别结果以置信度进行筛选。为此，SelectiveNet引入了一个拒绝选项，并围绕此设计损失函数，从而达到控制预测结果选择率的效果。拒绝选项的描述如下式所示：

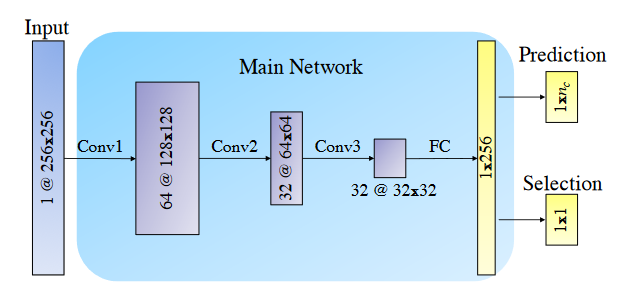
并在此基础上引入了覆盖率的概念，描述预测结果的选择情况：

并更新所用的普通CNN交叉熵损失函数

为了更进一步控制覆盖率为我们的预期值，进一步引入了惩罚项，强制使得在进行梯度下降时，覆盖率一定会高于设定值

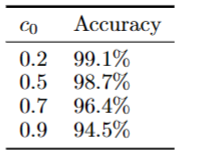


在此基础上建立的神经网络就可以达到所谓的筛选目的了。神经网络结构如下图所示



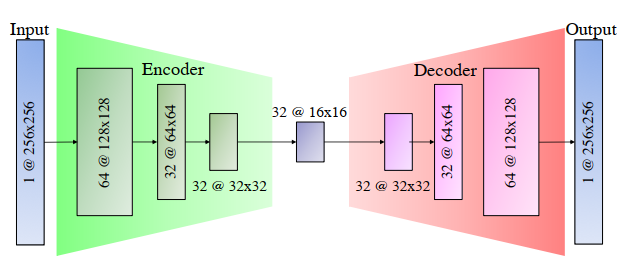
## 复现结果

我们在本地对实验模型成功进行了复现，并在原模型的基础上，在每一层卷积之间增加了BatchNorm，并使用了更加复杂的Selective Header，最终的实验结果如下，呈现了不同选择率下的结果：



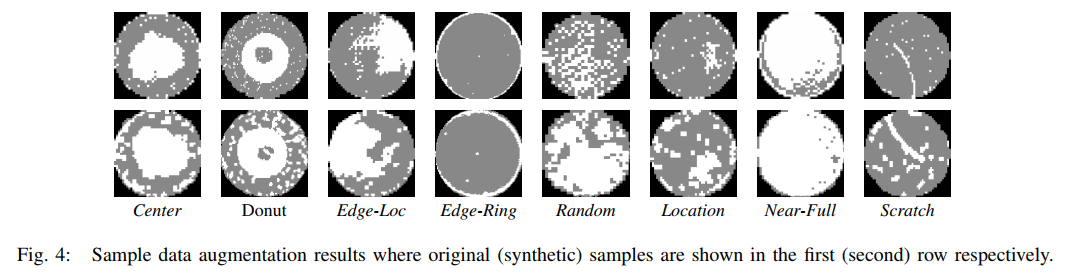
基本符合原论文结果，但我们也发现，该模型在测试集上泛化能力弱，在原论文中甚至没有给出在测试集上的结果。这也是我们在下一阶段中的优化方向

## 数据增强

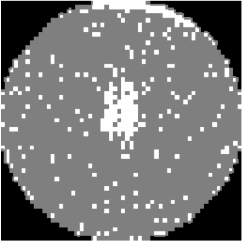


原论文中提到了对于判别器训练很重要的一点：数据集中不同类别数据的数量差异很大，呈现严重的不平衡，在总共54355张图片中，最少的类别Near-Full只有54张

而最多的种类None则有三万多张。如果不加以干预，会让模型的注意力集中在某几个类别上，而对其他的类别学习不够充分。因此我们按照原论文的数据增强方式实现了自编码器架构模型，以此生成相对真实的缺陷图片，以此来弥补原数据集中的不平衡。



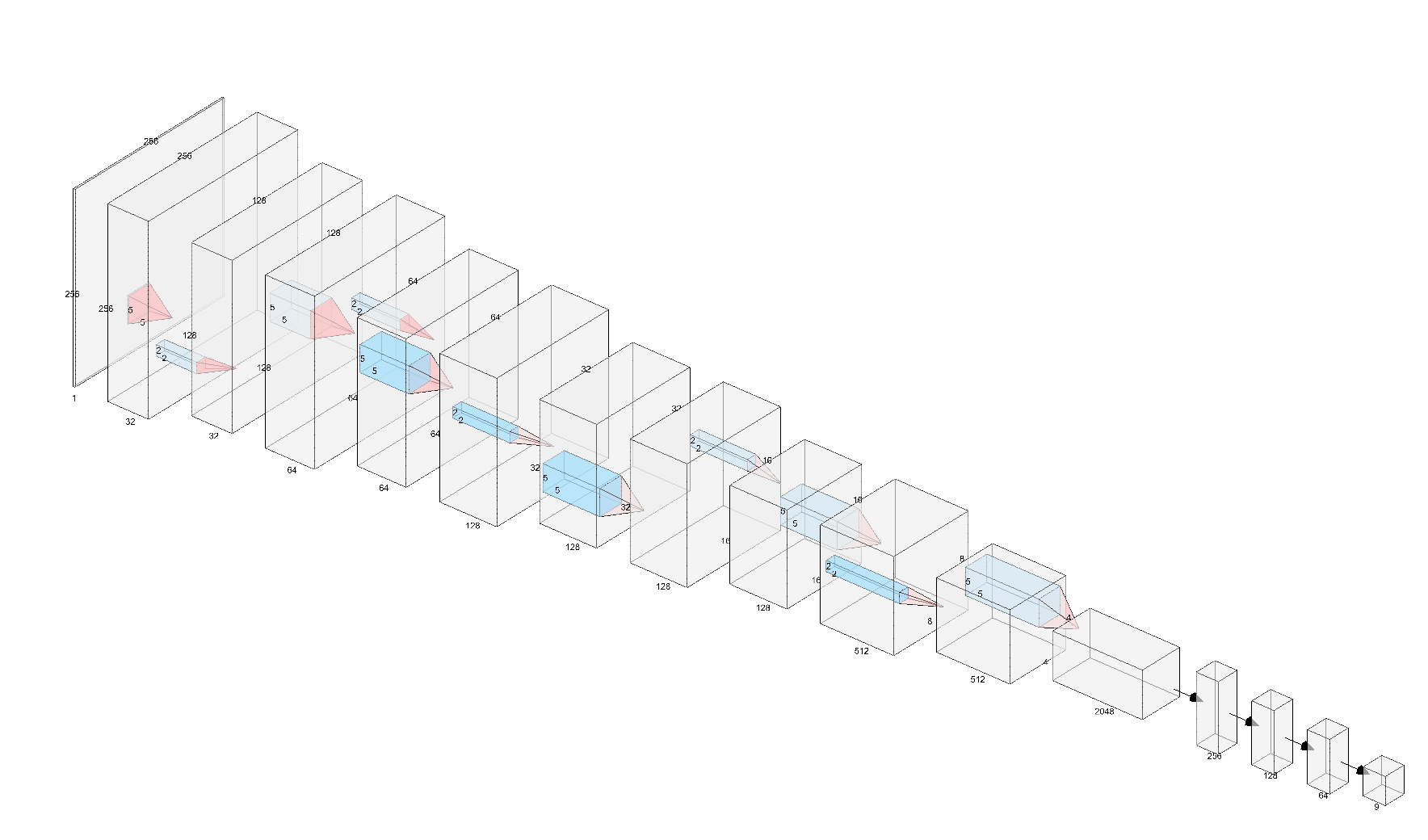
Auto-Encoder的模型架构为对称的Encoder与Decoder，在本论文中Encoder使用四次卷积，Decoder使用四次反卷积。中间的Dense Layer的size为32@16\*16。训练的目的是使得input与output尽可能相同，这样Encoder将学习到如何将WaferMap编码，Decoder将学到如何将编码解码为原图。

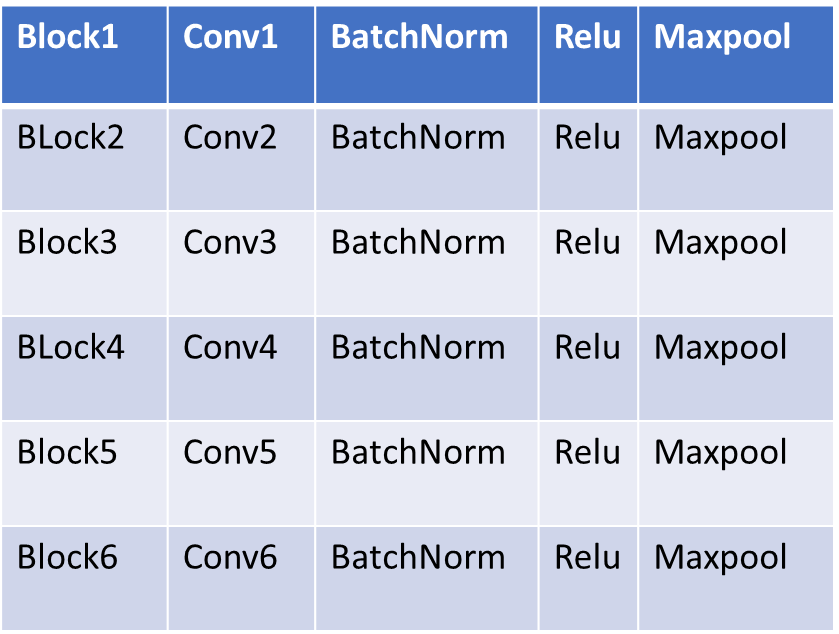


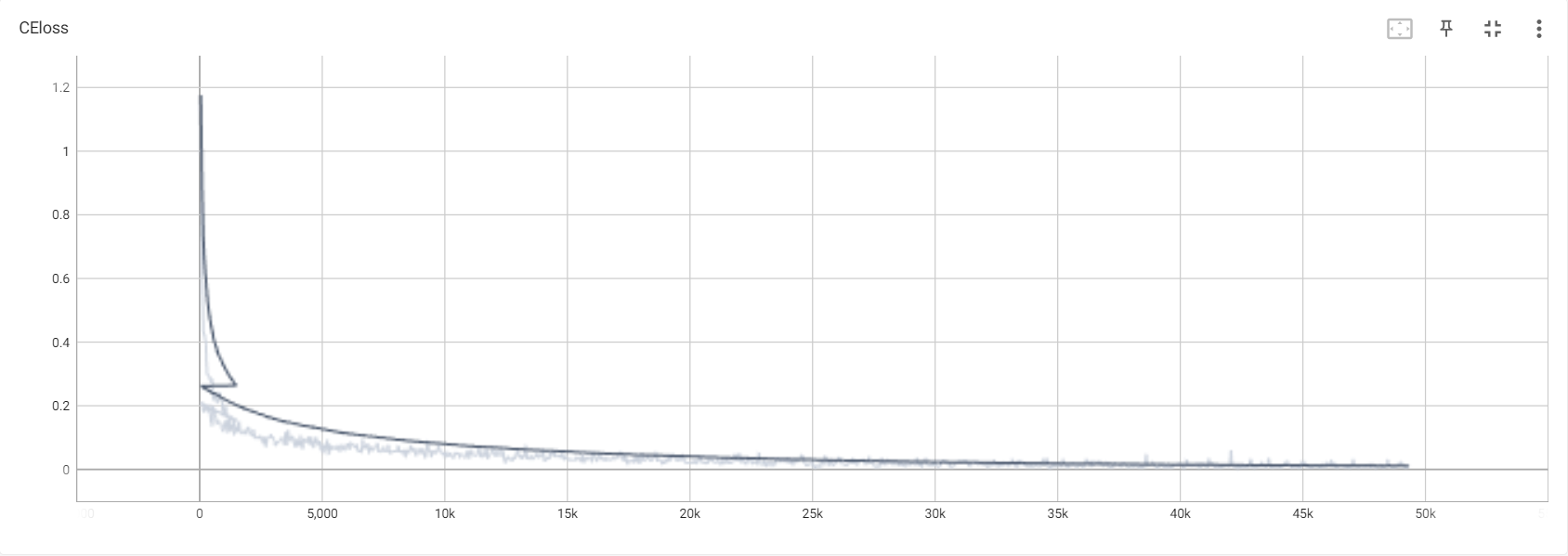
（左侧为center类的一张真实图片，右侧是基于该图片生成的结果）

在使用该模型进行数据生成时，对于某一类别的一张实际图片，我们先通过Encoder得到编码Latent，然后通过pytorch自带的正态噪声生成大小相同的噪声张量，将噪声与原Latent合并后输入到Decoder中，Decoder将对带噪Latent进行解码，解码后再对图片，旋转随机角度，得到一张新的训练图片。将每个类别都补充到5000张左右，便完成了数据增强的任务。

## 改进架构：Complex CNN

在实现原模型时，我们发现模型的欠拟合是比较严重的，在增强后的数据集上并没能达到原论文中的效果，准确率和覆盖率也都难以让人满意。因此，我们开始重新思考模型的架构。一方面，我们希望模型在卷积的不断深入中能够学习到更多的信息，因此我们将channel这一超参数设置为逐渐增加，而不是原作中的逐渐减少。另外，我们也思考了选择性学习的必要性。选择性学习本身是为了在覆盖率和准确率之间做出折中，也使得模型在遇见不属于这九类的真实样例时处理得更加谨慎。但是基于我们复现的原作模型，这种准确率和覆盖率之间的这种并不是令人满意的程度。因此，我们决定再实现一个更加复杂的CNN架构，并去掉选择性学习部分，实现100%覆盖测试集。

如上图所示，我们的卷积架构总共有6个Block，channel的数量努力呈现递增趋势：



（loss曲线）

训练过20个epoch后，留下了共49个检查点，让卷积网络加载每一个检查点，在118595张图片的测试集上进行测试。测试结果显示，表现最好的检查点是第32个，可以在训练时完全没有见过的共118595张图片的测试集上达到96.8952%的分类准确率。对比原作的选择性学习模型，该模型不仅coverage达到100%，并且在该条件下表现更好。

## 问题总结

1. 这项工作的意义：开发一个晶圆缺陷判别器本身的意义是减少工程师人工判断和标注的工作量，并通过判断晶圆缺陷的类别，方便更深层问题的解决。其次，这项工作中使用到了数据增强技术，可以为本身很少出现的缺陷类型生成大量的训练数据，从而解决训练数据不平衡给模型带来的一系列问题。这是一个很出色的自编码器的应用场景，作者的这一动机是我们认为本文比较出色的地方之一。
2. 对原模型的改进：我们对原模型进行了改进。在其启发下搭建了更为复杂的架构，使得覆盖率更高，准确率也更高。当模型欠拟合时，可以尝试扩大模型容量，可以对数据集进行更好的学习。
3. 对选择性学习的思考：原作提到的选择性学习的意义在于增加模型的置信程度，在准确率与覆盖率之间做出权衡。其优势在于，当有没有学习过的缺陷类型出现时，模型可以为工程师做出提示，转为人工检查，降低误判带来的损失。不过要合理权衡准确率与覆盖率，如果覆盖率太低，即便准确率很高，那么也没有实际使用的意义。

# 参考文献

1. M. B. Alawieh, D. Boning, and D. Z. Pan, “**Wafer Map Defect Patterns Classification using Deep Selective Learning**,” in *2020 57th ACM/IEEE Design Automation Conference (DAC)*, Jul. 2020, pp. 1–6. doi: 10.1109/DAC18072.2020.9218580.
2. P. P. Shinde, P. P. Pai, and S. P. Adiga, “**Wafer Defect Localization and Classification Using Deep Learning Techniques**,” IEEE Access, vol. 10, pp. 39969–39974, 2022, doi: [10.1109/ACCESS.2022.3166512](https://doi.org/10.1109/ACCESS.2022.3166512).
3. Y. Geifman and R. El-Yaniv, “SelectiveNet: A deep neural network with an integrated reject option,” in International Conference on Machine Learning (ICML), 2019.
4. “WM-811K wafer map,” https://www.kaggle.com/qingyi/ wm811k-wafer-map, accessed: 2019-07-30.
5. M.-J. Wu, J.-S. R. Jang, and J.-L. Chen, “Wafer map failure pattern recognition and similarity ranking for large-scale data sets,” IEEE Transactions on Semiconductor Manufacturing, vol. 28, no. 1, pp. 1–12, 2014.
6. S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, “Understanding data augmentation for classification: when to warp?” in 2016 international conference on digital image computing: techniques and applications (DICTA). IEEE, 2016, pp. 1–6.
7. S.-J. Jang, J.-S. Kim, T.-W. Kim, H.-J. Lee, and S. Ko, ‘‘A wafer map yield prediction based on machine learning for productivity enhancement,’’ IEEE Trans. Semicond. Manuf., vol. 32, no. 4, pp. 400–407, Nov. 2019.

# 仓库链接

[此处转到github仓库](https://github.com/ziqiao-w/Reproduce-Wafer-Map-Defect-Patterns-Classification-using-Deep-Selective-Learning)

1. 作者：王梓桥，计03班，学号： [↑](#footnote-ref-1)
2. 作者：吴成彰，计06班，学号：2019010333 [↑](#footnote-ref-2)