

本科生毕业设计（论文）开题报告

题　　目：基于对抗生成网络的闪存仿真器设计与实现

院　　系\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

计算机科学与技术学院

专业班级\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

计算机1604

姓　　名\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

孟嵩淼

学　　号\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

U201614613

指导教师\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

吴非

2020年3月**开题报告填写要求**

1. 开题报告主要内容：

1.课题来源、目的、意义。

2.国内外研究现况及发展趋势。

3.预计达到的目标、关键理论和技术、主要研究内容、完成课题的方案及主要措施。

4.课题研究进度安排。

5.主要参考文献。

1. 报告内容用小四号宋体字编辑，采用A4号纸双面打印，封面与封底采用浅蓝色封面纸（卡纸）打印。要求内容明确，语句通顺。
2. 指导教师评语、教研室（系、所）或开题报告答辩小组审核意见用蓝、黑钢笔手写或小四号宋体字编辑，签名必须手写。
3. 理、工、医类要求字数在3000字左右，文、管类要求字数在2000 字左右。
4. 开题报告应在第八学期第二周之前完成。

# 课题来源

该课题是本人在大四上学期在武汉光电国家研究中心实习时选择的课题，该课题由华中科技大学吴非带领研究。

对抗生成网络（GAN）是非监督式学习的一种方法，通过让两个神经网络相互博弈的方式进行学习。通常包含一个生成网络和一个判别网络，生成网络力图生成足以骗过判别网络的假数据，判别网络则需分辨出真实存在的数据和生成网络产生的假数据，这样通过不断地对抗学习，对于生成网络产生的假数据，判别网络最终既不能判定为假又不能判定为真。此时生成网络已可以产生足够以假乱真的假数据。GAN网络通常用于生成图片，视频或三维立体模型。本课题将使用生成对抗网络产生闪存块的错误信息，来代替从闪存中采集真实信息。

# 课题目的和意义

本课题将基于生成对抗网络设计并实现一个闪存仿真器，用以仿真闪存的性能、可靠性等行为，支持不同干扰特性定量组合的闪存数据生成。

研究该课题的直接意义在于，它可以对不同特征进行定量组合，生成可媲美测试数据的结果，大大降低3D闪存测试的时间成本。例如对于闪存寿命预测等问题需要针对闪存进行大量的测试来获得真实的块错误信息，进行这些测试需要花费大量时间，其次当一个块的P/E次数达到一定量时，该闪存块就会损坏。因此如果能使用仿真器生成的数据代替真实采集的数据，就可以节省采集数据所花费的时间。

# 国内外研究现况和趋势

本课题需要使用到GAN网络，该方法由伊恩·古德费洛等人于2014年提出，通过生成网络和判别网络相互博弈学习进行训练。生成网络接受一组随机取样值作为输入，其输出结果需要尽可能接近真实数据集中的样本。判别器则需要尽可能区分出真实数据和生成器产生的假数据。Mehdi Mirza后提出了CGAN网络，为GAN网络的一个变种，对于生成器，除了一组随机取样值外，还需添加一组值作为指定生成结果类型的标签；同样对于判别器除了以待判别数据作为输入外，也需要额外添加指定数据类型的标签值。CGAN网络训练至收敛后，可以通过输入网络的标签值来控制产生数据的类型，有别于GAN网络随机生成假数据；Alec Radford提出了DCGAN，为GAN网络的一个变种。在判别器的内部使用了卷积神经网络来代替全连接神经网络，在生成器内部则使用了转置卷积进行上采样来生成假数据，提高了生成器和判别器的性能。Martin Arjovsky提出了WGAN为GAN网络的一个变种，它去掉了判别器的Sigmoid层，去掉了损失函数的log层，每次更新判别器的参数后将其截断到一个区间[-c, c]（c为一固定常数），并且不再使用基于动量的优化算法，最终解决了GAN网络训练不稳定的问题，增加了生成样本的多样性。

目前已存在使用机器学习对闪存错误率进行预测的相关研究，基于2D MLC闪存芯片的实验数据,分别应用Elastic Net算法、BP神经网络算法及随机森林算法对2D MLC NAND Flash芯片进行错误率建模,并进行模型训练及模型结果的对比分析;为保证模型精度,分别基于3D TLC NAND Flash的短期驻留数据和长期驻留数据对3D芯片进行建模。

# 课题目标

本课题主要希望通过运用GAN网络及其变种，并对从闪存块采集的大量真实数据进行训练，最终得到可以产生媲美真实采样数据的生成器。

从个人的角度来说，希望通过毕业设计培养自己科学研究的能力，通过阅读论文获得所需使用技术的理论知识，通过查阅文献等方式学习技术工具的使用细节，最终构建一个解决目标问题的系统，并用技术手段去实现，最终完成课题目标，培养自己的学术能力和工程能力。

# 问题描述

块（Block）作为闪存的一个基本存储单元，它由2304个页组成，每个page又被划分为16个区域。块在多次P/E的过程中，这2304 × 16个位置均会产生一定量的错误。现对闪存进行测试，得到了一系列闪存记录文件，其中详细记录了块的在闪存中的位置，测试条件，以及块中每个位置测错误次数等信息。从这些记录文件中取得大量的块在不同P/E次数下的错误信息组织成真实数据集。通过GAN网络使用真实数据集进行训练，最终得到的生成器可以产生足以媲美真实数据的生成数据。

# 研究手段

首先是文献阅读，通过阅读相关论文学习GAN网络及其变种的理论知识，根据各种类型的GAN网络的特性来选择一种用于模型的构建。其次需要学习工程上所用的相关技术，构建机器学习模型使用开源框架Pytorch，因此需要掌握Python编程语言以及该框架的使用方法。对于大量的闪存测试数据进行分析，根据数据集特性进行模型的构建。构建完成模型后需对结果进行比较来选择合适的模型。

# 风险分析

风险来源一：可能欠缺研究该课题的知识或技能。课题中涉及较多的深度学习和统计学的知识，以及最终实现方案时需要掌握一些开源框架。通过查阅资料，文献来学习相关理论知识，并提前针对所需使用的编程语言和工具进行练习。

风险来源二：测试时可能欠缺高性能计算机，必要时导师会提供高性能计算机作为计算力的补充。

风险来源三：存在欠缺测试数据的可能性。本课题对模型进行训练需要使用大量的数据。

# 课题研究进度安排

表 1　课题研究进度安排表

|  |  |
| --- | --- |
| 月份 | 工作任务 |
| 2020年1月  ~2020年2月 | 接受课题，阅读文献，准备开发测试环境 |
| 设计系统的实现方案 |
| 2020年3月  ~2020年4月 | 准备开题答辩 |
| 实现方案 |
| 2020年5月 | 书写毕业论文 |
| 2020年6月 | 准备最终答辩 |

# 主要参考文献

[1] Bengio, Y., Mesnil, G., Dauphin, Y., and Rifai, S. (2013). Better mixing via deep representations. In ICML’2013.

[2] Bengio, Y., Thibodeau-Laufer, E., Alain, G., and Yosinski, J. (2014). Deep generative stochastic net- works trainable by backprop. In Proceedings of the 30th International Conference on Machine Learning (ICML’14).

[3] Frome, A., Corrado, G. S., Shlens, J., Bengio, S., Dean, J., Mikolov, T., et al. (2013). Devise: A deep visual-semantic embedding model. In Advances in Neural Information Processing Systems, pages 2121– 2129.

[4] Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In International Conference on Artificial Intelligence and Statistics, pages 315–323.

[5] Goodfellow, I., Mirza, M., Courville, A., and Bengio, Y. (2013a). Multi-prediction deep boltzmann ma- chines. In Advances in Neural Information Processing Systems, pages 548–556.

[6] Goodfellow, I. J., Warde-Farley, D., Mirza, M., Courville, A., and Bengio, Y. (2013b). Maxout networks. In ICML’2013.

[7] Goodfellow, I. J., Warde-Farley, D., Lamblin, P., Dumoulin, V., Mirza, M., Pascanu, R., Bergstra, J., Bastien, F., and Bengio, Y. (2013c). Pylearn2: a machine learning research library. arXiv preprint arXiv:1308.4214.

[8] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In NIPS’2014.

[9] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. Technical report, arXiv:1207.0580.

[10] Huiskes, M. J. and Lew, M. S. (2008). The mir flickr retrieval evaluation. In MIR ’08: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval, New York, NY, USA. ACM.

[11] Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y. (2009). What is the best multi-stage architecture for object recognition? In ICCV’09.

[12]郝梦琪. 基于机器学习的闪存错误率预测方法研究[D].哈尔滨工业大学,2019.

**华中科技大学本科生毕业设计（论文）开题报告评审表**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **姓名** |  | **学号** |  | **指导教师** |  |
| **院（系）专业** | |  | | | |
| **指导教师评语**   1. 学生前期表现情况。 2. 是否具备开始设计（论文）条件？是否同意开始设计（论文）？ 3. 不足及建议。 | | | | | |
| 指导教师（签名）：  年 月 日 | | | | | |
| **教研室（系、所）或开题报告答辩小组审核意见** | | | | | |
| 教研室（系、所）或开题报告答辩小组负责人（签名）：    年 月 日 | | | | | |