姓名：蒋睿兮 学号：191250063 文献综述方向：移动 工具复现方向：AI 工具复现论文（移动、AI）/选题标号（经典）：AI 《Counterexample-Guided Data Augmentation》

论文题目：反例引导式数据扩充

以反例为指导的数据扩增方法

要点：

扩增机器学习的反例数据集，反例是指分类错误示范的实例。

将所提出的框架与经典增强技术进行比较，以展示该框架的有效性。

背景：目前的机器学习算法生成的模型在可信度上面存在问题。让模型精度更高则需要更多的数据，于是对于机器学习来说，需要进行数据扩增来增加数据的数量。

数据扩增：数据扩增也叫数据增强，意思是在不实质性的增加数据的情况下，让有限的数据产生等价于更多数据的价值。

生成对抗网络：一种能够生成合成数据的特殊神经网络，来膨胀训练集

有一个叫错误表的新型数据结构。错误表对于提供有关反例的说明以及查找可能导致图像分类错误的重复模式非常有用。错误表分析还可以用于生成可能是反例的图像，从而有效地构建扩增集。错误表对于提供有关反例的解释和查找可能导致图像被错误分类的重复模式非常有用。

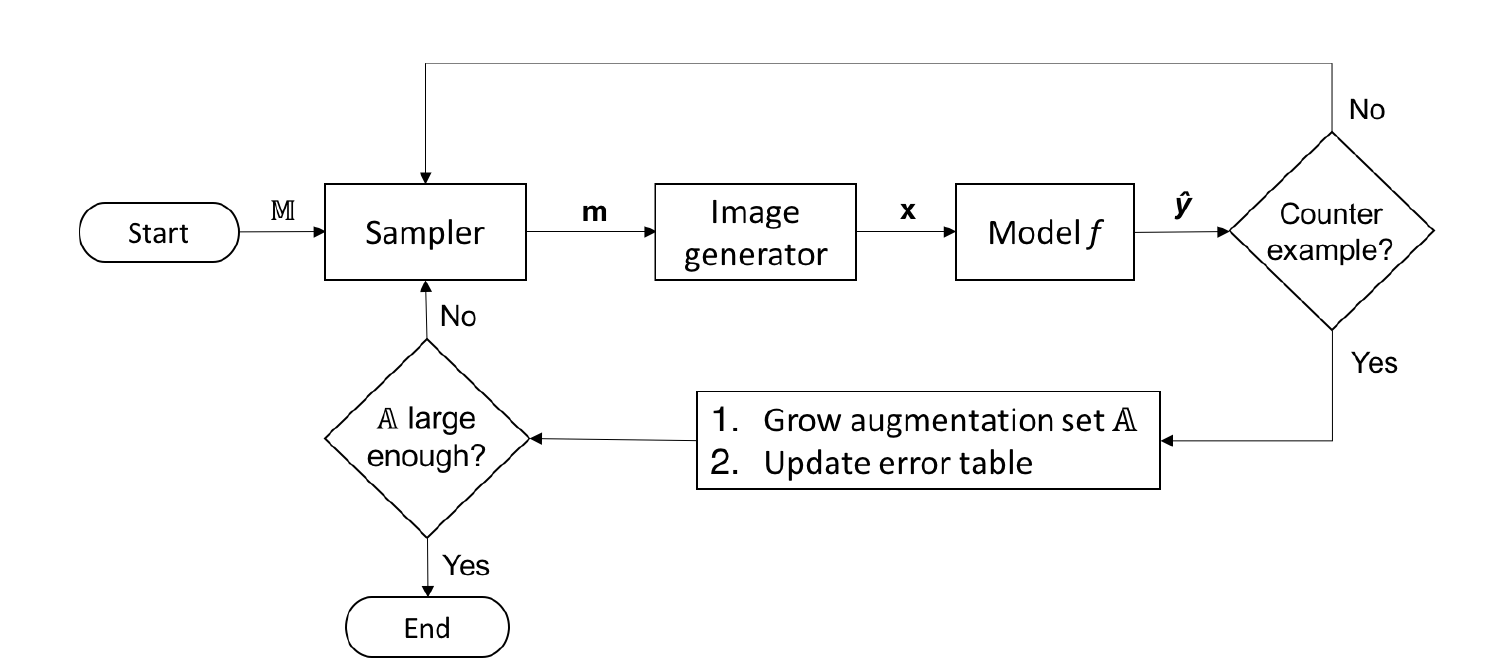
是一种数据扩增，数据扩增是指不实际增加原始数据，只是对原始数据做一些变换，从而创造出更多的数据。这样可以解决机器学习所需的数据不足的问题。

对抗网络：一种能够生成合成数据的特殊类型的神经网络。

主要思想是仅使用新的分类错误的示例来扩增训练集，而不是使用原始训练集中的修改后的图像。

卷积神经网络

训练集

反例数据的生成过程：

反例引导的增强方案。 M 表示修改空间，是一个采样的修改，x 是从 m 生成的图像，^y 是模型预测

简而言之就是通过M生成样本，然后判断样本是不是反例，当反例足够多了就停止生成。

反例扩增方案的核心是一个图像生成器，可渲染逼真的道路场景的合成图像。

抽样方法

低差异序列 在我们的实验中，我们使用Halton [Niederreiter， 1988]序列。

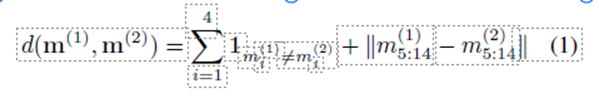
Halton序列

交叉熵采样

召回率 就是用你查出来的正确的数量除以所有正确的数量，可以跟准确率对比着记忆，

准确率是用你查出来的正确的数量除以所有的数量（包含正确和不正确的数量）。

定义了一个度量来衡量不同图片之间的区别



错误表

错误表分析可用于：

1.提供有关反例的说明，

2.生成反馈以采样新的反例。

错误表的特征分析 对于有序和无序特征的分析

主成分分析

奇异值分解

我们的方法可以看作是反例指导的归纳综合(CEGIS)的实例。

分析神经网络

一个用于分析和改进用于自动驾驶的 CNN 的平台。

主要成分：

图像生成器：渲染用于欺骗和重新训练 CNN 的合成图片

采样器：从合成图片空间采样配置

分析器：跟踪统计数据（例如，生成的图像、错误分类的特征、评估指标等）