反例引导式数据扩充

托马索·德罗西， 斯罗莫纳·戈什， 象屿岳，

库尔特·库策尔， 阿尔贝托·桑乔瓦尼-文森泰利， 桑吉特·塞希亚

加州大学伯克利分校

fdreossi j shromona.ghosh j xyyueg@berkeley.edu，

fkeutzer j alberto j [sseshiag@eecs.berkeley.edu](mailto:sseshiag@eecs.berkeley.edu)

**抽象**

我们提出了一个基于反例的用于增强机器学习数据集的新框架。反例是错误分类的示例，具有用于重新训练和改进模型的重要属性。我们框架的关键组件包括一个反例生成器，它生成由模型和错误表错误分类的数据项，这是一种新颖的数据结构，用于存储与错误分类有关的信息。错误表可用于解释模型的漏洞，并用于有效地生成用于增强的反例。通过对基于深度神经网络的自动驾驶中物体检测的案例研究，将所提出的框架与经典增强技术进行比较，展示了该框架的有效性。

**1 引言**

由机器学习算法产生的模型，特别是深度神经网络，正在部署在可信度是一个大问题的领域，从而产生了对更高准确性和保证的需求[Russell等人，2015;Seshia et al.， 2016]。然而，使用深度学习学习高精度模型受到对大量数据的需求的限制，甚至更进一步，受到劳动密集型标签的需求的限制。数据增强通过使用标签保留转换（即不改变标签的转换）来膨胀训练集，从而克服了数据的缺乏。传统的数据增强方案[van Dyk和Meng，2001;Simard等人， 2003;Cires ̧an et al.， 2011;Ciregan et al.， 2012;Krizhevsky等人，2012]涉及改变图像几何形状的几何变换（例如，旋转，缩放，裁剪或翻转）;和光度变换，这些变换会改变颜色通道。这些技术的功效最近已经得到证实（参见，例如，[Xu等人，2016;Wong et al.， 2016]）。传统的增强方案，如上述方法，将数据添加到训练集，希望提高模型的准确性，而不考虑模型已经学习了什么样的特征。最近，提出了一种复杂的数据增强技术[Liang等人，2017;Marchesi， 2017]，它使用生成对抗网络[Goodfellow等人，2014]，一种能够生成合成数据的特殊神经网络，来膨胀训练集。还有一些增强技术，例如硬阴性挖掘[Shrivastava等人，2016]，它用有针对性的阴性例子来膨胀训练集，以减少误报。在这项工作中，我们提出了一种新的增强方案，即反例引导的数据增强。主要思想是仅使用新的错误分类示例来扩充训练集，而不是使用来自原始训练集的修改图像。所提出的增强方案包括以下步骤：1）生成被模型错误分类的合成图像，即反例;2）将反例添加到训练集中;3） 在增强数据集上训练模型。可以重复这些步骤，直到达到所需的精度。请注意，我们的增强方案取决于生成错误分类图像的能力。出于这个原因，我们开发了一个图像生成器，它与采样器合作，产生作为模型输入的图像。图像的生成方式使得可以自动添加真实值标签。分类不正确的图像构成添加到训练集的增强集。除了图片之外，图像生成器还提供有关错误分类图像的信息，例如元素的配置，亮度，对比度等。此信息可用于查找反例中经常重复出现的特征。我们在我们称之为"错误表"的数据结构中收集有关反例的信息。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。误差表分析还可用于生成可能成为反例的图像，从而有效地构建增强集。

综上所述，这项工作的主要贡献是：

• 一种反例引导的数据扩充方法，其中只有错误分类的示例被迭代地添加到训练集中;

• 合成图像生成器，可渲染逼真的三重样本;

• 错误表，用于存储有关反例的信息，其分析提供解释并促进反例图像的生成。

我们通过分析不同反例数据增强采样方案，对卷积神经网络（CNN）进行对象检测，并将所提出的方法与经典数据实验方法进行了比较。我们的实验显示了使用反例驱动的方法与传统方法相比的好处。这种改进来自这样一个事实，即反例增强集包含模型无法从训练集中学习的信息，这是经典增强方案未考虑的事实。在我们的特殊情况下，我们使用由图像生成器生成的合成数据集。这可确保所有处理的数据都来自同一分布。

**概述**

图1总结了所提出的反例引导增强方案。该过程将修改空间M作为输入，M是我们图像生成器的可能配置的空间。空间M是基于领域知识构建的，是一个"语义修改"的空间;也就是说，每个修改都必须在使用机器学习模型的应用领域中具有意义。这使我们能够执行更有意义的数据，而不是简单地通过扰动输入向量（例如，对抗性地选择和修改图像中的少量像素值）来执行对抗性数据生成。

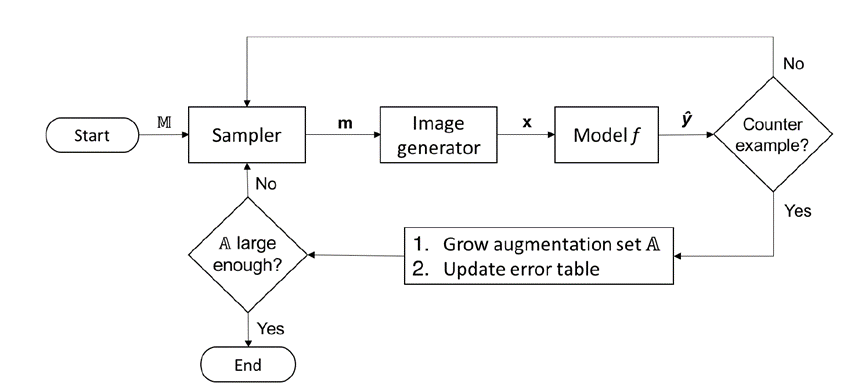


图1

在每个循环中，采样器从 M 中选择一个修饰项 m。样本由采样方法确定，该方法可能由预先计算的误差表偏倚，误差表是一种数据结构，用于存储有关模型错误分类的图像的信息。采样的修改由图像生成器呈现为图片 x。图像 x 作为返回预测 yˆ 的模型 f 的输入给出。然后，我们检查x是否是反例，即预测yˆ是错误的。我们将 x 添加到增强集 A 中，并将 x 的 infor-mation（例如 m，yˆ）存储在错误表中，采样器将在下一次迭代中使用。循环重复，直到增强集A足够大（或者M已经足够科学地覆盖了）。

此方案返回一个增强集，该集将用于重新训练处理的模型，以及一个错误表，其分析可识别反例之间的共同特征，并帮助采样器选择候选反例。

论文结构主要遵循图1的方案：第2节引入了一些符号;第3节描述了用于渲染合成图像的im-age生成器;第4节介绍了一些采样技术，可用于有效地对修改空间进行采样;第5节介绍了error表，并详细介绍了如何使用它们来提供关于反例的例证;第6节通过评估所提出的技术并比较反例引导的扩增方案和针对经典增强的不同调整方法，总结了paper。拟议框架的实施和报告的实验可在[https://github.com/dreossi/analyzeNN。](https://github.com/dreossi/analyzeNN)

**2 初步**

本节提供本文中使用的表示法。设 a 是一个向量，ai 是它的第 i 个元素，索引从 i = 1 开始，ai：j 是 a 从 i 到 j 的元素范围;A 是一个集合。X 是一组训练示例，x（i） 是数据集中的第 i 个示例，y（i） 是关联的标签。f ： A → B 是具有域 A 和范围 B 的模型（或函数）f。yˆ = f（x） 是模型 f 对输入 x 的预测。在对象检测上下文中，yˆ 对 f 为图像 x 预测的边界框、分数和类别进行编码。fX 是在 X 上训练的模型 f。设 B1 和 B2 为由 yˆ 编码的边界框。Uni on （IoU） 上的交集被定义为 IoU（AB1 ， AB2） = AB1 ∩ AB2 /AB1 ∪ AB2 ，其中 ABi 是 Bi 的面积，i ∈ {1， 2}。我们认为 Byˆ 是 By if IoU（Byˆ ， By） > 0.5 的检测。真阳性 tp 是正确的检测次数;误报 fp 是与任何基本实况框不匹配的预测框的数量;假阴性是未检测到的地面实况框的数量。

精度和召回率定义为 p（yˆ， y） = tp/（tp + fp） 和 r（yˆ， y） = tp/（tp + fn）。在这项工作中，如果 p（yˆ， y） 或 r（yˆ， y） 小于 0.75，则我们认为输入 x 被错误分类。设 T = {（x（1）， y（1））， . . . ， （x（m）， y（m））} 是具有 m 个示例的测试集。f 的 aver 精度和重调用 d ed 为 apf （T） = 1 r（f（x（i））， y（i））。我们使用平均 preci-i=1 sion 和召回率来测量模型的准确性，简洁地表示为 accf （T） = （apf （T），arf （T））。

**3 图像生成器**

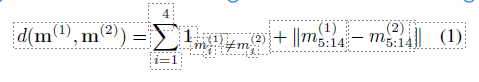
我们的反例增强方案的核心是一个图像生成器（类似于[Dreossi et al.， 2017a;Dreossi et al.， 2017b]），渲染道路场景的逼真合成图像。由于反例是由合成数据生成器生成的，因此我们完全了解所生成数据的地面实况标签。例如，在我们的例子中，当图像发生器将汽车放置在特定位置时，我们确切地知道它的位置和大小，因此相应地确定了地面实况边界框。在本节中，我们将介绍图像生成器的详细信息。

**3.1 修改空间**

图像生成器实现了生成函数γ：M → X，它将每个修改m∈M映射到γ（m）∈X的特征。例如，三维修改空间可以表征汽车x（横向）和z（远离）在道路上的位移和图像亮度。生成器可用于抽象和紧凑地表示高维图像空间的子集。

我们实现了基于14D修改空间的图像生成器，其尺寸决定了道路背景;汽车数量（一辆、两辆或三辆）及其在道路上的x和z位置;图片的亮度，清晰度，对比度和颜色。图 2 描述了我们的图像生成器渲染的一些图像。

我们可以在修改空间上定义一个度量来衡量不同图片的多样性。直观地说，如果具体化图像在视觉上是多样化的，则两种配置之间的距离很大，相反，如果具体化图像相似，则两种配置之间的距离很小。

以下是可以在我们的 14D 修改空间上定义的公制距离示例。设 m（1）、m（2） ∈ M 为修饰。距离定义为：

其中 1 条件为 1（如果条件为真），否则为 0，以及|| · || 是 L2 规范。距离计算了补间背景和汽车模型的差异，并添加了对应于 x 和 z 位置、亮度、锐度、对比度和图像颜色的点的欧氏距离。

图2描绘了三个图像及其修改m（1），m（2）和m（3）。为简洁起见，标题仅报告图像之间不同的尺寸，即背景、汽车模型和 x、z 位置。修改之间的距离为 d（m（1）， m（2）） = 0.48， d（m（1）， m（3）） = 2.0， d（m（2）， m（3）） = 2.48。请注意类似的图像，如图2（a）和（b）（相同的背景和汽车模型，略有不同的汽车位置），与图（a）和（c）等不同图像相比，距离（d（m（1），m（2）） = 0.48） 较小;或 （b） 和 （c） （不同的背景、车型和车辆位置），其距离为 d（m（1）， m（3）） = 2.0 且 d（m（2）， m（3）） = 2.48。

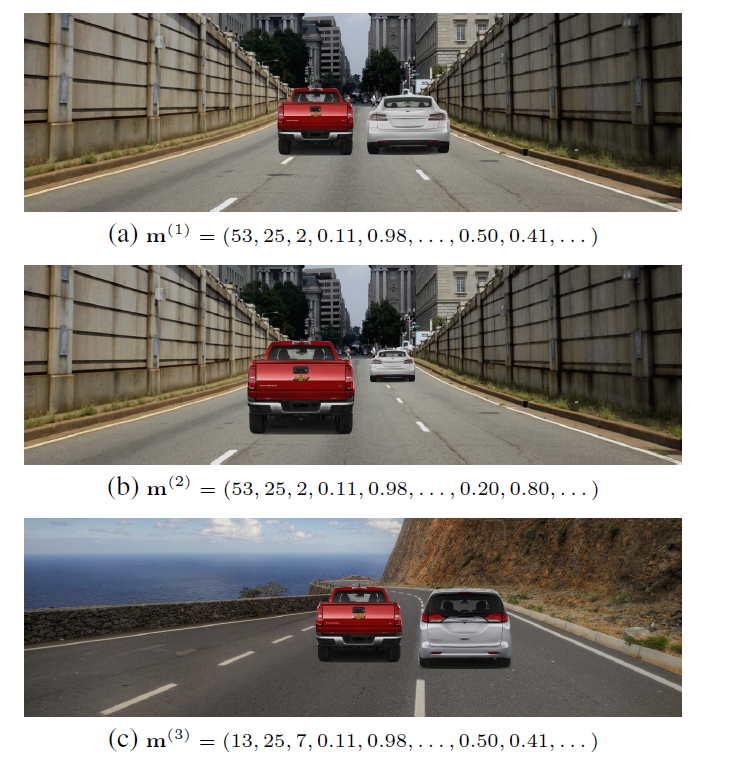


图2

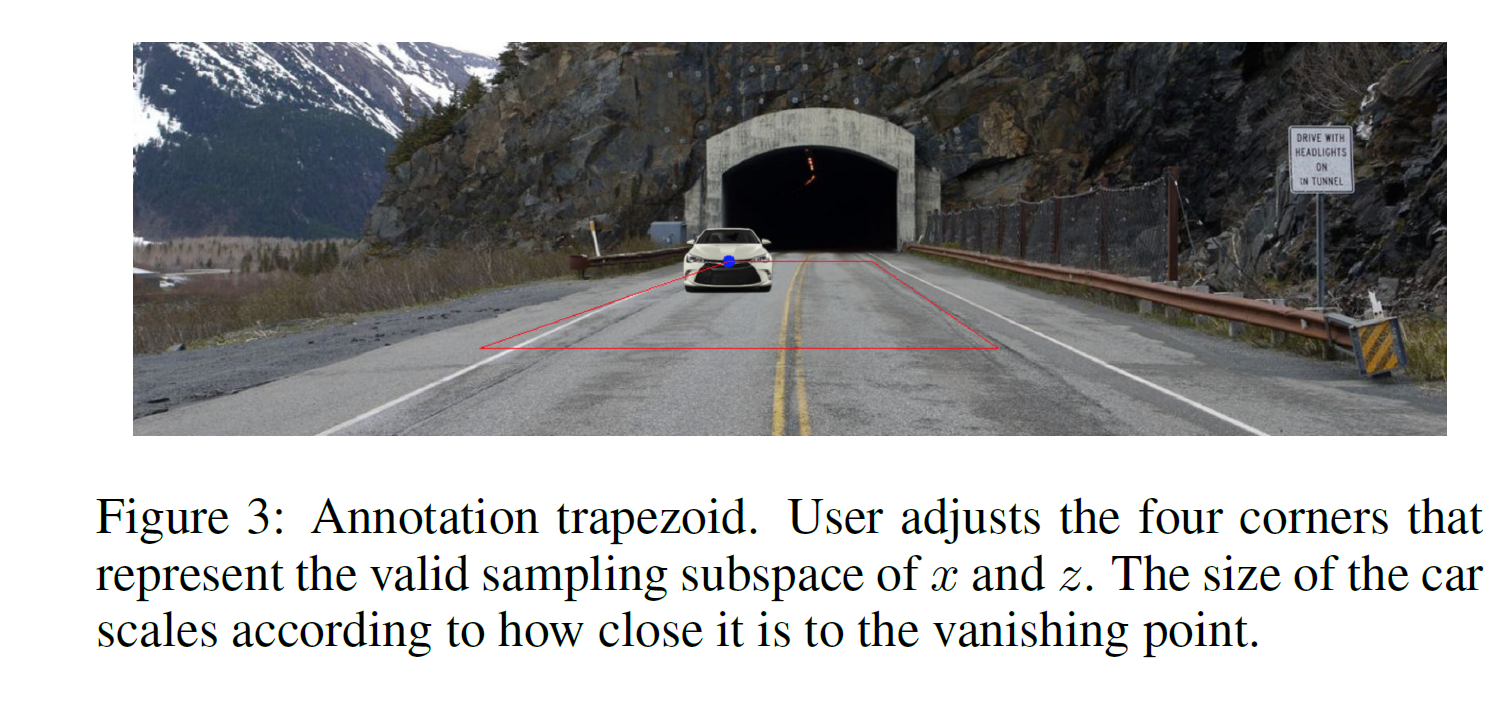
稍后，我们使用此指标来生成其元素确保一定多样性的集合。（见第6.1节）

**3.2 图片具体化**

修改修复后，我们的图片生成器将呈现相应的图像。具体化是通过叠加基本图像（如道路背景和车辆）并相应地根据修改指定的值调整图像参数（如亮度、颜色或对比度）来完成的。我们的图像生成器附带一个用作基本图像的地面和汽车模型数据库。我们的数据库由 35 个道路场景（例如，沙漠、森林或高速公路场景）和 36 个车型（例如，经济型、家庭型或跑车，从前视图和后视图）组成。数据库可以很容易地扩展或被用户替换。

**3.3 注释工具**

为了渲染逼真的图像，图片生成器必须将汽车放置在道路上并相应地缩放它们。为了便于将描述 x 和 z 位置的修改点转换为道路上汽车图像的正确叠加，我们为图像生成器配备了注释工具，该工具可用于指定道路上的采样区域和车辆的比例因子。对于特定道路，用户绘制一个梯形，指定允许图像发生器放置汽车的区域。用户还指定梯形底座上汽车图像的比例，即在距离观察者最远和最远的点处（见图3）。当在中间位置（即梯形内部）对点进行采样时，该工具会插值所提供的汽车刻度并确定给定点处的刻度。此外，图像生成器根据图像的视角叠加不同的车辆。图像生成器还会执行多项检查，以确保渲染的汽车可见。



图三

**4 采样方法**

采样器的目标是提供对修饰空间的良好覆盖，并识别其具体化导致反例的样品。

我们现在简要描述一些采样方法（simi-lar到[Dreossi et al.， 2017a;Dreossi et al.， 2017b]），我们将其集成到我们的框架中：

• 统一随机抽样：统一随机抽样可确保从M中抽样任何可能点的概率相等，从而保证了用于训练和测试的良好生成图像组合。虽然这是一种简单有效的技术，既适用于训练，也可用于测试，但它可能无法很好地覆盖修改空间;

• 低差异采样：低差异（或准随机）序列是 n 元组序列，它比不相关的随机点更均匀地填充 nD 空间。低差异序列有助于通过减少间隙和点聚类来覆盖盒子，从而确保样品空间的均匀覆盖。

• 交叉熵采样：交叉熵方法被开发为组合优化和重要性采样的通用蒙特卡罗方法。它是一种迭代采样技术，我们从给定的概率分布中采样，并通过最小化交叉熵来更新分布。

低差异序列的一些例子是Van der Corput，Halton [Halton，1960]或Sobol [Sobol，1976]序列。在我们的实验中，我们使用Halton [Niederreiter， 1988]序列。具有最佳覆盖率有两个主要优点：首先，我们增加了快速发现反例的机会，其次，反例集将具有很高的多样性;这意味着具体化的图像看起来会有所不同，因此模型将学习各种新功能。

**5 错误表**

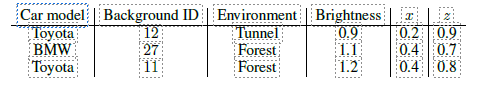
我们的增强方案的每次迭代都会产生一个反例，其中包含指向学习模型局限性的信息。最好提取与反例相关的模式，并使用此信息有效地生成新的反例。因此，我们定义了错误表，这些错误表是数据结构，其列由生成的图像中的重要特征组成。误差表分析适用于：

1.提供反例说明，以及

2.生成反馈以采样新的反例。

在第一种情况下，通过在反例中找到共同模式，我们向用户提供反馈，例如"该模型没有检测到在森林道路上驶离我们的白色汽车";在第二种情况下，我们可以将采样器偏向于更有可能导致反例的修改。

表 1：证明反例信息的错误表示例。第一行描述图 3。隐性无序特征：车型、环境;显式有序特征：亮度、x、z车坐标;显式无序功能：后台 ID。



表一

5.1 错误表功能

我们首先提供错误表支持的功能类型的详细信息。我们沿着两个维度对特征进行分类：

1.显式与隐式特征：显式特征是从修改空间（例如，x，z位置，亮度，对比度等）中采样的，而隐式特征是用户提供的生成图像的各个方面（例如，汽车模型，背景场景等）。

2.有序与无序特征：某些特征具有具有明确定义的总排序（例如，清晰度）的域，而其他特征则没有排序的概念（例如，汽车型号，背景场景的标识符等）。

隐式和显式功能集是互斥的。通常，隐式特征更具描述性，并表征生成的图像。这些对于提供反馈以解释分类器的漏洞非常有用。虽然隐式特征是无序的，但显式特征可以是有序的，也可以是无序的。错误表行是用于错误分类的特征的实现。

表 1 是一个说明性错误表。该表包括汽车模型和环境场景（隐式无序特征）、亮度、x、z 汽车坐标（显式有序特征）和背景 ID（显式无序特征）。表 1 的第一行实际上是指图 3。我们的框架生成的实际错误表大于选项卡 1。例如，它们包括我们的14D修改空间（见第3.1节）和车辆数量，车辆方向，背景中的主导颜色等功能。

给定一个填充了反例的错误表，我们希望对其进行分析以提供反馈，并利用此反馈对新图像进行采样。

5.2 特征分析

一种朴素的分析技术是平等对待所有特征，并在错误表的每一列中搜索最常出现的元素。但是，这假设特征之间没有相关性，而事实往往并非如此。相反，我们为有序和无序特征开发单独的分析技术。在下文中，我们将讨论如何最好地捕获两个集合之间的相关性：

有序特征：由于这些特征是有序的，因此有意义的分析技术是在发生大多数伪造样本的特征空间中查找分辨。这与使用主成分分析（PCA [will et al.， 1987]）的模型阶数减少非常相似。具体来说，我们对第一个主成分感兴趣，它是对应于矩阵的奇异值分解（SVD [will et al.， 1987]）中最大奇异值的奇异向量，该矩阵由所有有序特征的样本组成。我们可以使用奇异向量来查找模型相对于有序特征的敏感程度。如果向量中与特征对应的值较小，则意味着模型对该特征的变化不可靠，即该特征的变化会影响错误分类。或者，对应于奇异向量中较大值的特征充当"不关心"，即通过修复所有其他特征，模型无论该特征的值如何，都会对图像进行错误分类;

无序特征：由于这些特征是无序的，因此它们的值并不重要。我们可以从中收集到的最有意义的信息是最常一起出现的特征子集。为了正确地捕捉到这一点，我们必须探索所有可能的子集，这是一个组合问题。当无序要素的空间很大时，这被证明是有问题的。克服此问题的一种方法是限制要探索的最大子集的大小。

我们在一组500个反例上进行了实验。制定的功能包括每辆车的x和z位置;以及整个图像的亮度，对比度，清晰度和颜色。显式功能包括有序功能以及所有可能的汽车和背景的离散集。隐含的功能包括汽车的颜色，背景的颜色，汽车的方向等细节。显式有序特征上的PCA显示对应于第一辆车的x位置（0.74），亮度（0.45）和对比度（0.44）的高值。我们可以得出结论，该模型对这些有序特征的变化并不可靠。对于无序特征，森林道路与一辆白色汽车（后部朝向摄像机，另一辆车面向摄像机）的组合出现了13次。这为反例中反复出现的元素提供了解释，特别是"该模型没有检测到在森林道路上驶离我们的白色汽车"。

5.3 使用反馈进行采样

我们可以利用误差表分析提供的反馈来指导抽样以进行后续训练。请注意，我们只能从显式功能中采样：

• 来自有序特征的反馈：有序特征是显式特征的子集，已经告诉我们在采样过程中哪些特征需要更多变化。例如，在 5.2节 的示例中，我们的采样器必须优先对第一辆车的不同 x 位置进行采样，然后是亮度，最后是其他有序特征之间的对比度;

• 来自无序特征的反馈：设 Suf = Sef ∪ Sif 是分析重新转换的大多数已发生无序特征的子集，其中 Sef 和 Sif 分别是显式和隐式特征的互斥集。Sef的信息可以直接合并到采样器中。Sif 提供的信息需要一些推理，因为隐式特征不是直接采样的。然而，它们与图像的特定元素（例如，背景或车辆）相关联。

表2：增强技术的比较。报告精度（顶部）和召回率（底部）。用采样方法T生成的TT集;在X上训练的fXT模型f使用技术T∈{S，R，H，C，D，M};S：标准，R：均匀随机，H：Halton低差异，C：交叉熵，D：具有距离约束的均匀随机，M：所有方法的混合。

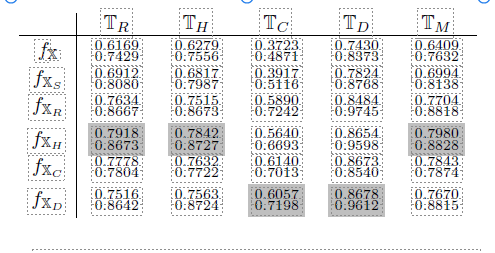


表2

我们可以使用图像生成器库和错误表来识别库中哪些元素的组件Sif 对应于要素值，并相应地设置要素值。例如，在 5.2节的示例中，对无序显式特征的分析表明，桥梁道路与特斯拉、梅赛德斯和马自达的组合最常被错误分类。我们使用这些信息通过改变亮度和对比度来生成更多具有这种特定组合的图像。

第6.3节展示了这种技术如何产生可用于再训练的更大比例的反例。

**6 实验评估**

在本节中，我们将展示如何使用所提出的技术来增强训练集并提高所考虑模型的准确性。我们将尝试不同的采样方法，将反例引导增强与经典增强进行比较，迭代几个增强周期，最后展示误差表如何成为分析模型的有用工具。关于所提议的框架的实施情况和所报告的实验，可在 https://github.com/dreossi/analyzeNN 中查阅。

在所有实验中，我们分析了squezeDet [Wu et al.， 2016]，这是一种用于自动驾驶的CNN实时对象检测器。所有模型都经过了65个迭代的训练。

原始训练和测试集 X 和 T 分别包含我们的图像生成器随机生成的 1500 张和 750 张图片。初始精度 accfX （T） =

（0.9847， 0.9843） 相对较高（见表 3）。但是，我们将能够生成大到以下的反例集：fX 的精度下降的 T。表中突出显示的条目显示了最佳性能。报告的值是五个不同实验的平均值。

6.1 增强方法比较

作为第一个实验，我们使用不同的采样技术运行反例增强方案（见第4节）。具体而言，我们考虑了均匀随机抽样、低差异荷顿序列、交叉熵抽样和对采样点具有多样性约束的均匀随机抽样。对于后者，我们采用第3.1节中的距离，我们要求反例的修改必须彼此之间至少相距0.5。

对于每种采样方法，我们生成 1500 个反例，其中一半注入到原始训练中集合 X 和一半用作测试集。让 R、H、C、D 对均匀随机、荷顿、交叉熵和多样性（即具有距离约束的随机性）采样方法进行分解。设 T ∈ {R， H， C， D} 作为采样技术。XT是X的增强，TT是一个测试集，两者都使用T生成。为了实验的完整性，我们还定义了测试集 TM，其中包含由所有 R、H、C、D 采样方法生成的反例的相等组合。

表2报告了使用各种增强集训练的模型的准确性，这些增强集在使用不同采样技术生成的反例的测试集上进行评估。第一行报告了在原始训练集 X 上训练的模型 fX 的准确性。请注意，尽管模型在原始测试集（accfX （T） = （0.9847， 0.9843）） 上的准确性很高，但我们还是能够从模型性能较差的 X 和 T 的相同分布中生成多个测试集。

我们考虑的第一个扩增是标准的，即我们使用imgaug1更改X的图像，imgaug1是一个用于图像增强的Python库。我们通过随机裁剪每侧10 − 20%，水平翻转，概率为60%，并在σ ∈[0.0，3.0]的情况下应用高斯模糊，从而增强了X中50%的图像。标准增强提高了每个测试集的准确性。各种测试集的平均精度和召回率分别为4.91%和4.46%（见第1行表2）。

接下来，我们使用反例引导方案（均匀随机、低差异 Halton、交叉熵和带距离约束的随机）来增强原始训练集 X，并在各种测试集上测试重新训练的模型。均匀随机的平均精度和召回率改进分别为 14.43% 和 14.56%，低差异 Halton 的平均精度和召回率分别为 16.05% 和 14.57%、交叉熵 16.11% 和 6.18%，距离约束为 14.95% 和 14.26%。首先，请注意，与经典的增强方法相比，使用反例引导增强方法的原始模型的准确性提高更大。其次，在所提出的技术中，交叉熵在精度方面具有最高的改进，但低差异性方法在精度和召回率方面往往比其他方法平均表现更好。这是因为低差异序列确保了样本上比其他技术更多的多样性，从而产生不同的图片，模型可以从中学习新特征或增强弱特征。

为原始模型生成反例时，均匀随机采样平均为 30s、Halton 92s 和约束为 55s 的均匀随机采样。这显示了模型精度的时间和增益之间的权衡。最大化增强集的多样性（以及随后的准确性提高）需要更多的迭代。

6.2 增强循环

对于此实验，我们仅考虑统一随机抽样方法，并在多个增强循环中增量扩充训练集。这个实验的目的是了解模型是否过度拟合反例，看看是否有可能达到饱和点，即我们无法生成反例的模型。我们还有兴趣研究注入反例的数量与模型的准确性之间的关系。

考虑第 i 个增强周期。对于每个增强轮次，我们通过考虑具有最高平均精度的模型f并重调用来生成一组反例。给定 X[i]，我们的分析工具生成一组反例 C[i]。我们将 C[i] 分成两半 C[i] 和 C[i]。我们使用 C[i] 来增强原始训练集 X[i] 和 C[i] 作为测试集。具体来说，增强训练集 X[ri+1] = X[ri] ∪ C[i] 是通过向 X[i] 添加错误分类的 C[i] 图像而获得的。r， r 是错误分类的图像与原始训练示例的比率。例如，|X0.08|= |十|+ 0.08 ∗| X|，|十|是 X 的基数。我们考虑比率 0.08， 0.17， 0.35， 0.50。我们针对每个测试集评估每个模型。

选项卡 3 显示了三个增强周期的精度。对于每个模型，该表显示了相对于原始测试集 T 和错误分类图像的测试集的平均精度和召回率。第一个循环的生成大约需要6个小时，第二个循环需要14个小时，第三个循环需要26个小时。50多个小时后，我们停止了第四个周期。这表明，为经过多次增强训练的模型生成反例变得越来越困难。反例生成的计算难度随着周期数的增加，是机器学习模型中增加保证的非正式经验衡量标准。

请注意，对于每个周期，我们的增强功能都会提高模型相对于测试集的准确性。更有趣的是，原始测试集上的模型精度不会降低，而是实际上会随着时间的推移而提高（至少对于所选的增强比率）。

6.3 误差表引导采样

在最后一个实验评估中，我们使用误差表来分析通过均匀随机抽样为fX生成的反例。我们分析了有序和无序特征（见第5.2节）。PCA对有序特征的分析揭示了以下相关值：清晰度0.28，对比度0.33，亮度0.44和x位置0.77。这告诉我们，模型对图像变化比对其元素的配置更敏感。对无序特征的出现次数统计显示，错误分类中出现次数最多的前三名车型是白色保时捷，黄色Corvette和浅绿色菲亚特。有趣的是，与流行的汽车相比，所有这些车型都有不寻常的设计。最常出现的前三个背景场景是森林中的一座狭窄的桥梁，一个室内停车场和一个市中心的城市环境。所有这些场景的特征都是高密度的细节，导致错误阳性。使用收集到的信息，我们将采样器空间缩小到由误差表分析确定的修改空间的子集。反例属-tor能够产生329个错误分类与10k迭代，而纯均匀随机抽样有103个，Halton有287个，均匀随机有96个具有距离约束。

最后，我们在训练集 XE 上重新训练了 f，其中包括使用误差表分析生成的 250 张图像。被玷污的精度是

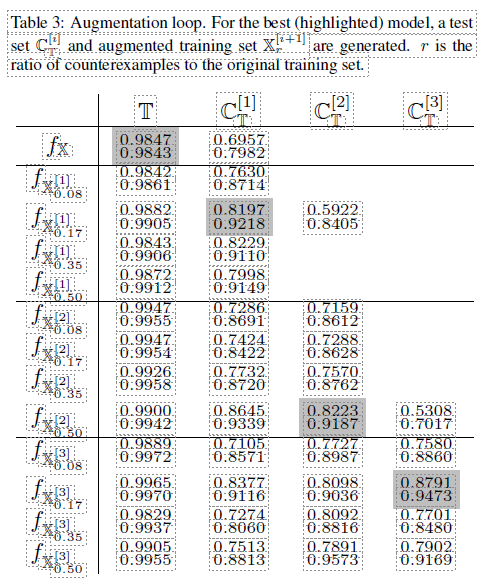
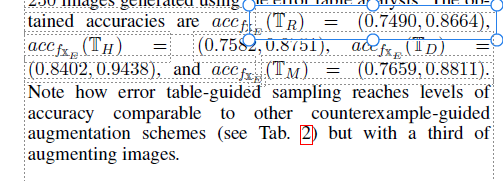


表3



请注意误差表引导采样如何达到与其他反例引导的增强方案相当的准确度水平（见表 2），但具有三分之一的增强图像。

**7 结论**

在本文中，我们提出了一种使用反例增强机器学习（ML）数据集的技术。我们生成的反例是综合生成的数据项，这些数据项被机器学习模型错误分类。由于这些项目是通过算法合成的，因此它们的基本事实标签也是自动生成的。我们展示了如何使用误差表来有效地指导增强过程。训练深度神经网络的结果表明，我们的增强技术在图像分类方面的表现优于标准增强方法。此外，当我们迭代增强循环时，在计算上很难找到反例。我们还表明，误差表可以通过较小的数据增强有效地提高准确性。

我们注意到，我们提出的方法也可以倾向于使用来自"系统级"分析和验证的反例，其中人们分析了周围环境中整个系统（例如，自动驾驶功能）的正确性[Dreossi等人，2017a]。使用这种"语义反例"执行数据增强是未来工作的一个有趣的方向[Dreossi等人，2018]。

我们的方法可以被视为反例引导归纳合成（CEGIS）的一个实例，CEGIS是程序合成中的一种com-mon范式[Solar-Lezama等人，2006;Alur et al.， 2013]。在我们的例子中，正在合成的程序是ML模型。CEGIS本身是预言机引导的归纳合成（OGIS）的一个特例[Jha和Seshia，2017]。对于未来的工作，探索使用除反例生成预言机以外的预言机来扩充数据集，并将我们的反例引导数据增强技术与其他预言机引导的数据扩充方法进行比较，这将是很有趣的。

最后，在这项工作中，我们决定完全依赖模拟的综合数据，以确保训练、测试和反例集来自同一数据源。将我们的增强方法扩展到现实世界的数据会很有趣;例如，驾驶过程中道路场景的图像。为此，人们需要使用诸如领域适应或迁移学习之类的技术[Tobin等人，2017]，这些技术可以将合成生成的数据适应现实世界。

确认

这项工作部分得到了NSF赠款1545126，1646208和1739816，协议号为FA8750-16-C0043的DARPA BRASS计划，DARPA保证自治计划，iCyPhy中心和Berkeley Deep Drive的支持。我們感謝 NVIDIA Corporation 透過捐贈用於這項研究的 Titan Xp GPU 的支持。Hisahiro （Isaac） Ito关于交叉熵采样的建议得到了感谢。