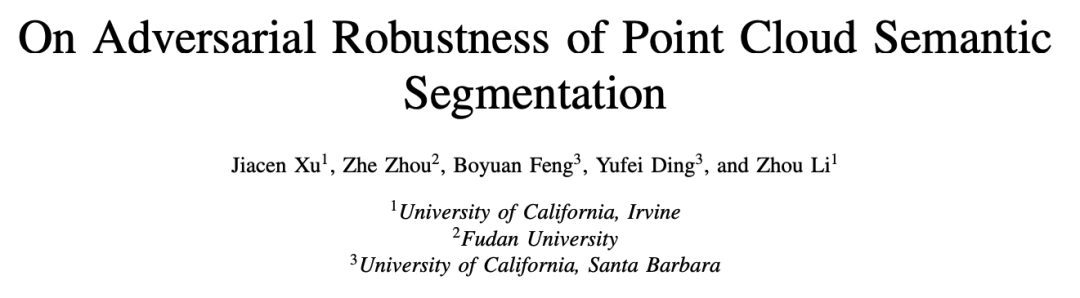
G.O.S.S.I.P 阅读推荐 2023-06-29

2023-06-29 21:26

今天要给大家推荐的论文是来自加州大学尔湾分校数据驱动安全和隐私实验室，复旦大学以及加州大学圣巴巴拉分校的研究者们关于其最新研究On Adversarial Robustness of Point Cloud Semantic Segmentation的投稿。目前由于自动驾驶、无人快递、无人超市等应用的兴起，针对三维点云语义分割模型（Point Cloud Semantic Segmentation, PCSS）的研究也成为一个热门方向。但是关于这类模型的安全性并没有一个系统的研究，本工作主要针对三维点云语义分割模型的健壮性和安全性做进一步的探索，分析不同数据类型对模型安全的影响。基于不同的攻击目标，文章设计了目标隐藏攻击和性能降级攻击；基于攻击类型，设计了范数受限攻击以及范数不受限攻击。目前该研究已被DSN 2023录用。在DSN大会期间，他们将于6月30号11:00-12:30现场分享（葡萄牙波尔多的Bolsa Palace，Dourada会议室），欢迎关注。



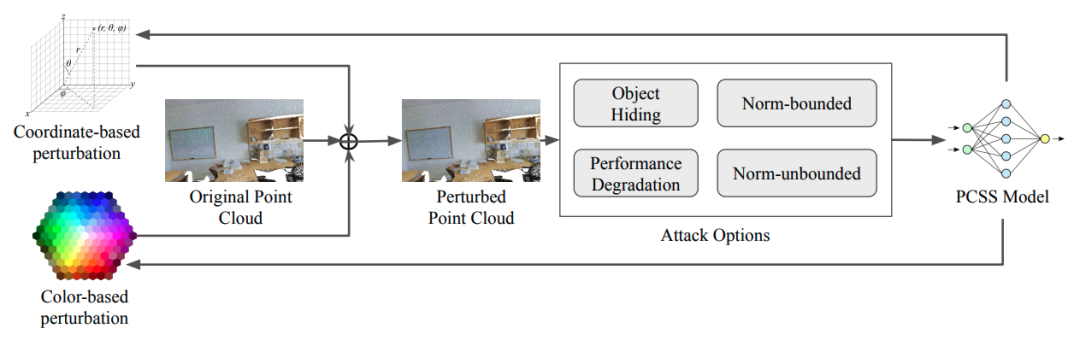
在自动系统领域，准确性和健壮性是非常关键的因素，决定了整个系统在现实环境是否试用。因此，大量相关研究孕育而生。针对准确性，研究者运用深度学习模型或图神经网络来处理点云数据。而针对健壮性，研究者通过借鉴图像领域的对抗机器学习算法来扰动三维点云数据以此误导模型。通过分析，作者发现这些方法只关注了物体识别应用，而关于语义分割以及PCSS模型的健壮性研究却比较少。文章因此展开针对三维点云语义分析模型健壮性的研究，并发现PCSS与普通图像的本质区别：

数据意义不同：相比较普通图片在固定像素的颜色进行扰动，三维点云的颜色和每个点的位置都可以移动，使得攻击者搜索的扰动范围更大，如何高效的扰动成为重要的问题。

数据预处理方式不同：相比较对图片颜色正规化的处理，不同PCSS模型会设计不同的方式进行点云分割、点云取样等操作。这些操作都会对模型的健壮性效果产生影响。

攻击目标不同：图片分类的目标是使得图片整体的唯一标签被错误地分类，而PCSS模型的攻击目标是需要考虑整个点云中所有点的标签。

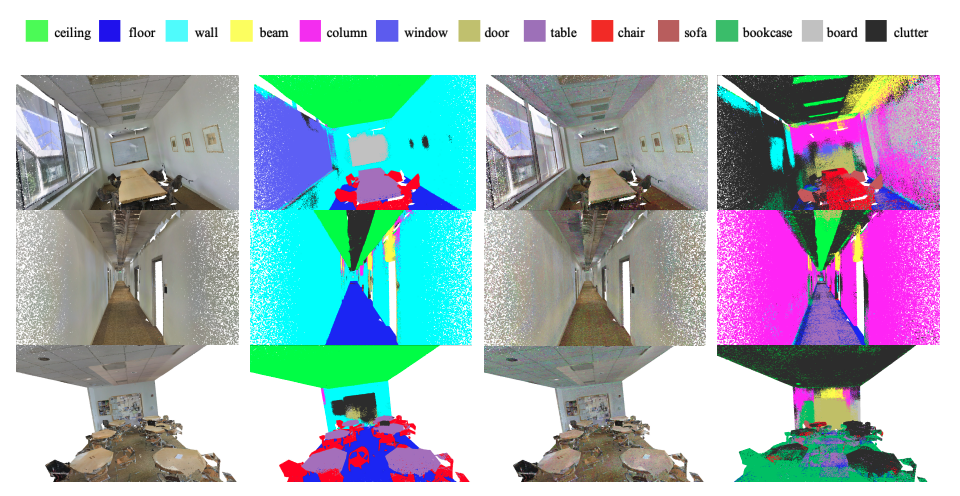
在不同场景下，文章设计了不同形式的攻击算法。基于不同的攻击目标，本文设计了目标隐藏攻击和性能降级攻击；基于攻击类型，设计了范数不受限攻击以及范数不受限攻击。攻击流程图如下所示：

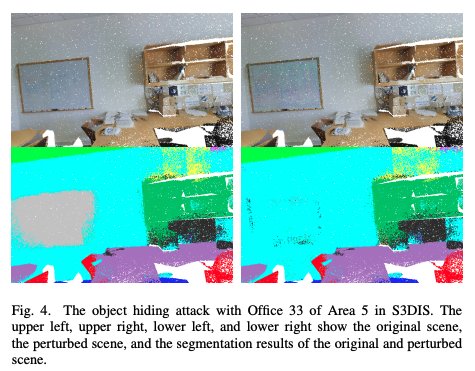


在原始点云的基础上，分别针对位置或颜色进行扰动，把扰动过的点云基于不同的攻击要求和PCSS模型计算Loss并更新扰动。在不断迭代优化的过程中，完善扰动最终达到误导PCSS模型的目的。

具体来说，目标隐藏攻击将针对某一类物体，例如桌子进行扰动使得PCSS将其分类成墙体；而性能降级攻击则不考虑错误的种类，只考虑使分类结果出错。范数受限攻击将在攻击过程中主动限定扰动的范数，如果攻击过程中范数超过限定大小就进行切割。范数不受限攻击则不主动限定范数，通过在Loss Function中加入计算范数的一项使得在扰动优化过程中控制范数大小。四种攻击的具体细节可以参考文章。

本文将所设计的攻击算法分别在室内数据库S3DIS以及室外的数据库Semantic3D上进行实验，并在三种不同PCSS模型ResGCN-28, PointNet++以及RandLA-Net上测试健壮性。以下是一些物体隐藏攻击和整体表现降级攻击的攻击结果展示





实验结果表明：

相比较位置，颜色的扰动更容易误导PCSS模型。

范数受限攻击和范数不受限攻击都有显著的攻击效果。并且在一些较困难的点云中，范数不受限攻击效果要优于范数受限攻击。

在目标隐藏攻击中，不同的目标类型攻击难度不同。例如，当希望使得一些点云被分类成墙面时，原始点云是门、书架、白板等会更容易攻击成功。

在这项工作中，作者首次对三维点云语义分割（PCSS）的对抗攻击进行了比较性研究。作者系统地设计了目标隐藏攻击和性能降级攻击，并基于范数受限攻击和范数不受限攻击开发了多种攻击方法。除了所有现有对抗攻击都利用的点云坐标位置之外，作者还考虑对点云颜色等特征进行扰动。作者在室内数据集S3IDS和室外数据集Semantic3D上对这些攻击组合通过测量范数，来评估每种攻击对原始点云的影响。总体而言，作者发现所有经过测试的PCSS模型在对抗扰动下都容易受到攻击，尤其是应用于颜色特征的范数不受限攻击。作者希望通过这项研究，能够加强对PCSS模型的健壮性做出更多努力。

项目网站：https://c0ldstudy.github.io/posts/PC\_attack  
论文下载：https://c0ldstudy.github.io/commons/papers/dsn2023\_PointCloud.pdf  
文章代码：https://github.com/C0ldstudy/PointSecGuard

作者介绍：

徐嘉涔，加州大学尔湾分校电子与计算机工程博士三年级，师从李洲教授。本硕毕业于上海交通大学网络空间安全学院。研究方向包括数据驱动的计算机安全，对抗机器学习以及图神经网络在网络空间安全上的应用。  
个人主页：https://c0ldstudy.github.io/about/