

## Google 流体标注：机器学习助力探索性界面研发，大幅提升图像标注速度

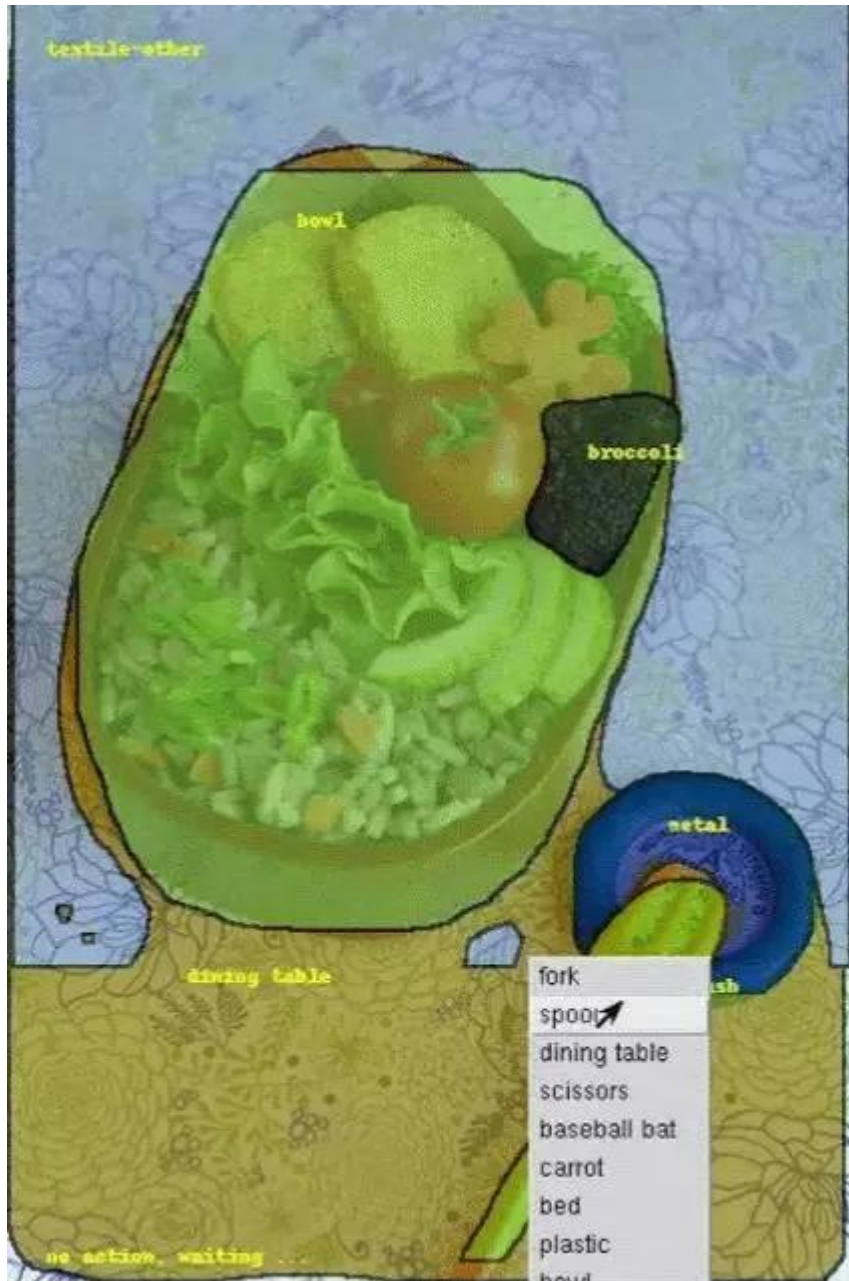
对于基于深度学习的现代计算机视觉模型（例如由 TensorFlow Object Detection API 实现的模型）而言，其性能取决于能否使用日益扩大的标注训练数据集（例如 Open Images）开展训练。然而，对优质训练数据的获取正迅速成为制约计算机视觉发展的主要瓶颈。对于在自动驾驶、机器人和图像搜索等应用中执行的逐像素预测任务（例如语义分割）而言，这一点体现得尤为明显。事实上，传统的手动标注工具需要标注者仔细点击物体的边界，以此勾勒出图像中的每个物体，这个过程枯燥乏味：标注 COCO+Stuff 数据集中的一个图像需时 19 分钟，而标注整个数据集则需花费 53000 多个小时！



COCO 数据集的图像示例（左）及其逐像素语义标注结果（右）。图像来源：Florida Memory，原始图像

在 “《流体标注：用于完整图像标注的人机协作界面》(Fluid Annotation: A Human-Machine Collaboration Interface for Full Image Annotation)” 一文中，我们探讨了一种以机器学习为技术支持的界面，以标注类别标签并在图像中勾勒每个物体的轮廓和背景区域，此举可将数据集的标注速度提升 3 倍。此论文将在 2018 年 ACM 多媒体会议的 “勇敢新理念 (Brave New Ideas)” 环节展示。

流体标注是从某个强大语义分割模型的输出开始，而人类标注者可以使用自然用户界面，并通过机器辅助编辑操作来修改此模型。标注者可通过我们的界面选择要修正的内容和修正顺序，进而以高效的方式集中精力去攻克机器尚未了解的方面。



对 COCO 数据集的图像应用流体标注界面的可视化效果。图像来源：gamene，原始图像

更确切地说，在标注图像之前，我们首先会通过一个预训练的语义分割模型 (Mask-RCNN) 来处理图像。这大约会生成 1000 个图像片段，且每个片段都带有相应的类别标签和置信度分数。我们使用置信度最高的片段初始化要向标注者呈现的标签。然后，标注者就可以：(1) 从机器生成的最终候选名单中选择标签，以更改现有片段的标签。(2) 添加片段以弥补缺失的物体。机器会识别出最有可能预先生成的片段，而标注者可以滚动查看，并选择最佳片段。(3) 移除现有片段。(4) 调整重叠片段的深浅顺序。如要进一步了解此界面，请试用演示版（仅限桌面设备）。



使用传统手动标注工具（中间列）和流体标注（右列）对三张 COCO 数据集图像进行标注的效果比较。虽然使用手动标注工具标注出的物体边界通常更为精确，但造成标注差异的首要原因在于人类标注者往往对于物体所属的确切类别存有分歧。图像来源：sneaka，原始图像（上）；Dan Hurt，原始图像（中）；Melodie Mesiano，原始图像（下）

流体标注标志着我们朝着更加轻松快速的图像标注之路迈出了探索性的第一步。在未来的研究中，我们将致力于改进对物体边界的标注，利用更多机器智能以提升界面速度，最终实现界面扩展，使之能够处理之前未曾见过的类别，满足其对于高效数据收集的迫切需求。

## 致谢

此项研究是与 Misha Andriluka 协作完成。特别感谢 Christine Sugrue 在创建流体标注演示版时所做的努力。我们还要感谢 Anna Ukhanova 和 Damien Henry 提供的宝贵意见。