

数据标注平台	
名称	链接
Mechanical Turk	https://www.mturk.com/
Figure-eight	https://www.figure-eight.com/
Mighty AI	被Uber收购，不开放了
数据堂	http://bz.datatang.com/
百度众测	https://test.baidu.com/
阿里众包	https://newjob.taobao.com/
京东微工	https://weigong.jd.com/

数据标注开源工具

名称	简介	运行平台	标注形式	导出数据格式	链接
LabelImg	著名的图像标注工具	Windows\Linux\Mac	矩形	XML 格式	https://github.com/tzutalin/labelImg
LabelMe	著名的图形界面标注工具,能标注图像和视频	Windows\Linux\Mac	多边形、矩形、圆形、多段线、线段、点	VOC 和COCO 格式	https://github.com/wkentaro/labelme
RectLabel	图像标注	Mac	多边形、矩形、多段线、线段、点	YOLO、KITTI、COCO1 与CSV 格式	https://github.com/ryouchinsa/Rectlabel-support
VOTT	微软发布的基于Web 方式本地部署的标注工具,能标注图像和视频	Windows\Linux\Mac	多边形、矩形、点	TFRecord、CSV、VoTT 格式	https://github.com/microsoft/VoTT
LabelBox	适用于大型项目的标注工具,基于Web、能标注图像、视频和文本		多边形、矩形、线、点、嵌套分类	JSON 格式	https://github.com/Labelbox/Labelbox
VIA	VGG(Visual Geometry Group)的图像标注工具,也支持视频和音频标注		矩形、圆、椭圆、多边形、点和线	JSON 格式	https://github.com/LakorTi/Via
COCO UI	用于标注 COCO 数据集的工具,基于Web 方式		矩形、多边形、点和线	COCO 格式	http://cocodataset.org/#home
Vatic	Vatic 是一个带有目标跟踪的视频标注工具,适合目标检测任务	Linux		VOC 格式	https://github.com/cvondrick/vatic
BRAT	基于Web 的文本标注工具,主要用于对文本的结构化标注	Linux		ANN 格式	https://github.com/nlplab/brat
DeepDive	处理非结构化文本的标注工具	Linux		NLP 格式	https://github.com/HazyResearch/deepdive
Praat	语音标注工具	Windows\Unix\Linux\Mac		JSON 格式	https://github.com/praat/praat
精灵标注助手	多功能标注工具	Windows\Linux\Mac	矩形、多边形和曲线	XML 格式	http://www.jinglingbiaozhu.com/

数据标注数据集

类别	数据集名称	用途	大小	链接	开放
图像标注数据集	ImageNet	图像分类、定位、检测	1TB	http://www.image-net.org/about-stats	是
	COCO	图像识别、分割和图像语义	40G	http://cocodataset.org/#home	是
	PASCAL VOC	图像分类、定位、检测	2GB	http://host.robots.ox.ac.uk/pascal/VOC/	是
	Open Image	图像分类、定位、检测	1.5GB	https://github.com/openimages/dataset	是
	Flickr30k	图片描述	30MB	http://shannon.cs.illinois.edu/DenotationGraph/	是
视频标注数据集	Youtube-8M	理解和识别视频内容	1PB	https://research.google.com/youtube8m/	受限
	Kinetics	动作理解和识别	1.5TB	https://deepmind.com/research/open-source/	是
	AVA	人类动作识别		https://research.google.com/ava/	是
	UCF101	视频分类、动作识别	6.5GB		是
文本标注数据集	Yelp	文本情感分析	2.66GB	https://www.yelp.com/dataset/challenge	是
	IMDB	文本情感分析	80.2MB	http://ai.stanford.edu/~amaas/data/sentiment/	是
	Multi-Domain Sentiment	文本情感分析	52MB	https://www.cs.jhu.edu/~mdredze/datasets/sentiment/	是
	Sentiment 140	文本情感分析	80MB	http://help.sentiment140.com/	是
语音标注数据集	LibriSpeech	训练声学模型	60GB		是
	AudioSet	声学事件检测	80 MB	https://research.google.com/audioset/	是
	FMA	语言识别	1000G	https://github.com/mdeff/fma	是
	VoxCeleb	语音识别、情绪识别	150MB		是