**ActivityNet dataset：**

1.2版本：

100 activity classes

4,819 training videos (7,151 instances)

2,383 validation videos (3,582 instances)

2,480 testing videos (labels withheld)

1.3 版本：

200 activity classes

10,024 training videos (15,410 instances)

4,926 validation videos (7,654 instances)

5,044 testing videos (labels withheld)

标记文件格式：

database

{

5n7NCViB5TU: {

annotations: [

{

label: "Discus throw",

segment: [24.25018, 38.08036]

},

{

label: "Discus throw",

segment: [97.00073, 106.284]

}

],

duration: 121.44,

resolution: "320x240",

subset: "training",

url: "https://www.youtube.com/watch?v=5n7NCViB5TU"

}

}

taxonomy

[

{

parentName: "Playing sports",

nodeName: "Discus throw",

nodeId: 92,

parentId: 64

},

{

parentName: "Participating in Sports, Exercise, or Recreation",

nodeName: "Playing sports",

nodeId: 64,

parentId: 17

},

{

parentName: "Sports, Exercise, and Recreation",

nodeName: "Participating in Sports, Exercise, or Recreation",

nodeId: 17,

parentId: 5

},

{

parentName: "Root",

nodeName: "Sports, Exercise, and Recreation",

nodeId: 5,

parentId: 0

},

{

parentName: null,

nodeName: "Root",

nodeId: 0,

parentId: null

}

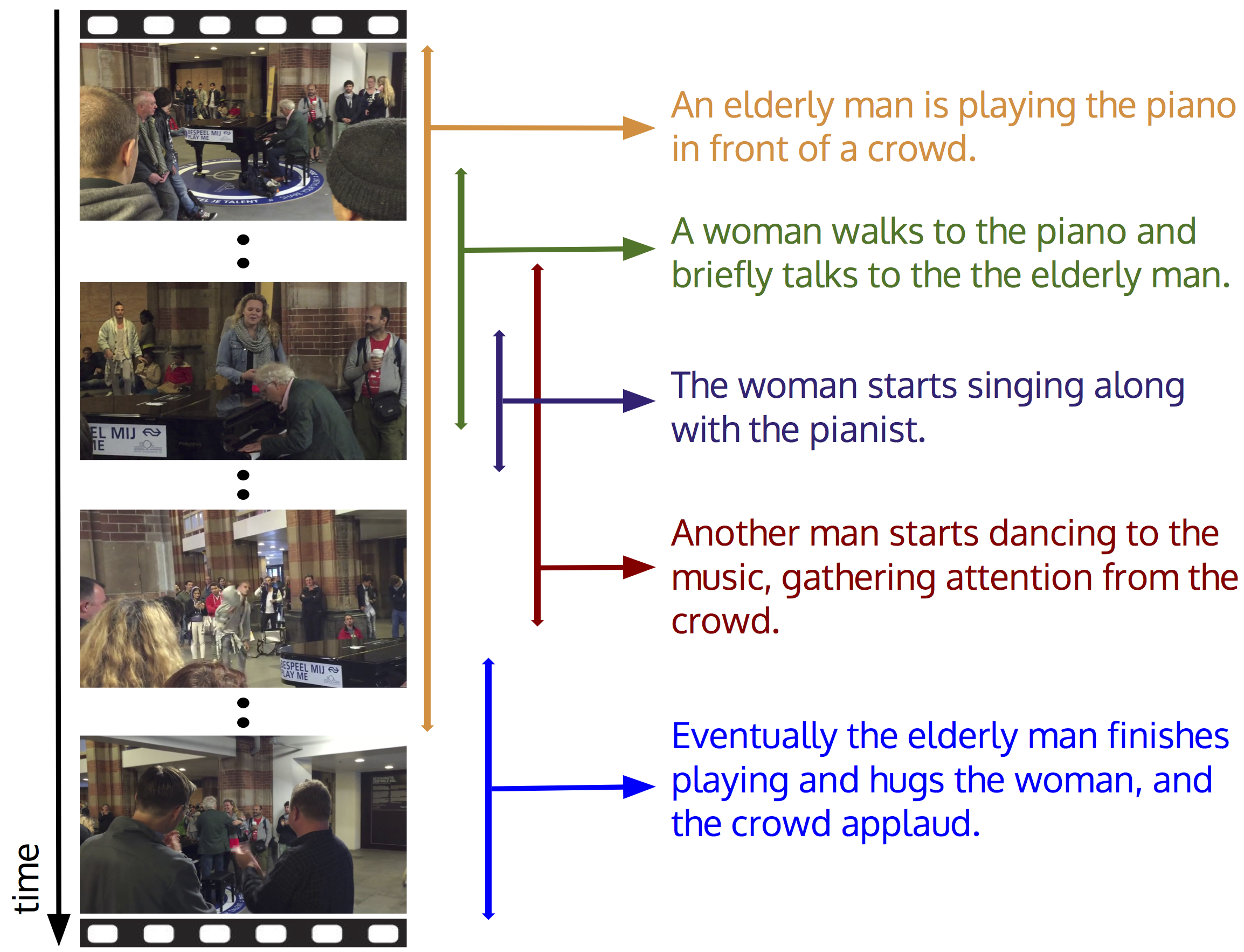
]

version

The key field "version" identifies which version of ActivityNet is contained in the annotation file. It is intended to keep track over different releases.

**ActivityNet Captions dataset：**

该数据集将视频与按时间顺序标记的句子描述建立联系。每个句子只覆盖视频中的一段，描述该段中出现的多个事件，这些事件可能持续很长时间也可能持续较短时间还可能并发。平均下来，每个视频包含3.65个标记好的句子描述，总共有20k个视频，100k个句子描述。每个视频的句子描述数量基本遵循正态分布，视频越长，句子描述数量越多。而每个句子描述的平均长度为13.48个单词，也是基本遵循正态分布。



**VidVRD dataset：**

该数据集基于ILSVRC2016-VID(Object detection from video task)的训练集与验证集构建的。而ILSVRC2016-VID数据集是人们为这些视频里出现的object标记bounding box而来的，总共有30个object categories。从ILSVRC2016-VID中精心挑选1000个视频，又额外增加频繁出现在visual relation的object categories，人工对额外的5个object进行类别和bounding box trajectories标注，又对每个视频分段进行predicate和visual relation instance的标注，下面是VidVRD的数据集统计信息：

