



(12) 发明专利申请

(10) 申请公布号 CN 112509110 A

(43) 申请公布日 2021.03.16

(21) 申请号 202011492370.3

(22) 申请日 2020.12.16

(71) 申请人 清华大学

地址 100084 北京市海淀区清华园1号

(72) 发明人 刘彬彬 朱纪洪 欧阳波 于帆

叶梓轩

(74) 专利代理机构 北京三聚阳光知识产权代理

有限公司 11250

代理人 张建纲

(51) Int.Cl.

G06T 15/50 (2011.01)

G06T 15/30 (2011.01)

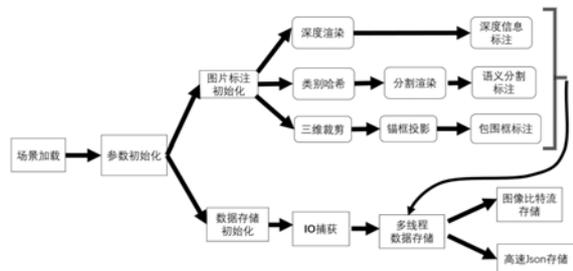
权利要求书2页 说明书5页 附图1页

(54) 发明名称

一种陆地对抗智能体的图像数据集自动采取与标注框架

(57) 摘要

本发明公开一种陆地对抗智能体的图像数据集自动采取与标注框架,该框架主要包括:参数初始化部件,负责读取配置文件,并设置框架的基本参数;图像标注部件,图像标注部件由深度信息标注,语义分割标注,物体包围框标注三个小部件组成,分别负责标注图像的深度信息,语义分割信息和图像中物体的包围框信息;数据存储部件,数据存储部件负责将图像标注部件的标注信息存储到磁盘。本发明所提的框架,能够针对陆地对抗的不同环境来建立合适的虚拟场景,然后实时并且精确地获得该陆地对抗场景下的图像数据集,利用得到的数据集可以训练陆地对抗智能体,提升其对抗素质和能力。



1. 一种陆地对抗智能体的图像数据集自动采取与标注框架,其特征在於,包括:自动实时标注图像深度信息的方法,自动实时标注图像的语义分割标签的方法,自动实时标注图像中物体包围框的方法,图像与图像标注自动实时存储的方法。

2. 根据权利要求1所述的框架,其特征在於,所述自动实时标注图像深度信息的方法包括以下步骤:

步骤S1,从场景中得到各个物体到相机的距离信息d;

步骤S2,将距离信息编码为灰度值;

步骤S3,Unity3D利用步骤S2得到的灰度值进行渲染得到的场景图片即为深度信息标注图。

3. 根据权利要求1所述的框架,其特征在於,所述自动实时标注图像的语义分割标签的方法包括以下步骤:

步骤S1,给不同类别的物体设置好不同的标签;

步骤S2,通过哈希映射的方式,将不同的标签编码为不同的颜色;

步骤S3,将物体渲染的颜色参数值设置为步骤S2所得到的颜色值;

步骤S4,Unity3D渲染后的图片就是图像语义分割的标签真值。

4. 根据权利要求1所述的框架,其特征在於,所述自动实时标注图像中物体包围框的方法包括以下步骤:

步骤S1,获取物体在三维场景中的三维包围框和物体的类别标签与类别名称;

步骤S2,根据物体间的遮挡关系,利用二分法对步骤S1所得的三维包围框进行裁剪;

步骤S3,利用坐标变换关系式将三维包围框的八个顶点投影到相机的二维屏幕上,得到屏幕上的8个点,记为 D_{1-8} ,每个点 D_i 的坐标为 (x_i, y_i) ,根据 D_{1-8} 可以得到二维屏幕上的包围框 $Rect(X_{min}, X_{max}, Y_{min}, Y_{max})$,然后根据如下式子得到二维包围框的标注真值;

$$\begin{cases} C_{id} \\ X_c = X_{min} + (X_{max} - X_{min})/2 \\ Y_c = Y_{min} + (Y_{max} - Y_{min})/2 \\ W = X_{max} - X_{min} \\ H = Y_{max} - Y_{min} \end{cases}$$

其中, C_{id} 为包围框所属物体的类别标签;

所述步骤S2进一步包括以下步骤:

步骤S21,对于步骤S1所得的三维包围框的每个面,如果面上的四个顶点均被遮挡,就反复执行十次步骤S22,如果面上的四个顶点没有完全被遮挡,则不做处理;

步骤S22,记当前所处理的面为 F_i ,与它相对的面为 F_j , F_i 与 F_j 的距离为d,判断 F_i 的四个顶点是否都被遮挡,如果是,就将 F_i 朝靠近 F_j 的方向移动d/2,如果否,则将 F_i 朝远离 F_j 的方向移动d/2;

步骤S3所述的坐标变换关系式为:

$$\begin{pmatrix} x_{camera} \\ y_{camera} \\ z_{camera} \\ w_{camera} \end{pmatrix} = \begin{bmatrix} \frac{n}{r} & 0 & 0 & 0 \\ 0 & \frac{n}{t} & 0 & 0 \\ 0 & 0 & \frac{-(f+n)}{f-n} & \frac{-2fn}{f-n} \\ 0 & 0 & -1 & 0 \end{bmatrix} \begin{pmatrix} x_{world} \\ y_{world} \\ z_{world} \\ w_{world} \end{pmatrix}$$

其中n和f分别为相机视锥的近平面距离和远平面距离,而t和r可以由相机的垂直视野角fovY及投影屏幕的宽高比aspect得出,具体关系由如下式子给出:

$$\begin{cases} \tan \frac{fovY}{2} = \frac{t}{|n|} \\ aspect = \frac{r}{t} \end{cases}$$

5. 根据权利要求1所述的框架,其特征在于,所述图像与图像标注的自动存储的方法包括以下步骤:

步骤S1,设置自动存储图像与图像标注的间隔时间;

步骤S2,根据步骤S1的间隔时间,将渲染后的图片,深度信息图片和语义分割标注图片存储到磁盘,然后将所标注的物体包围框,物体类别标签和类别名称统一转为Json格式存储到磁盘。

一种陆地对抗智能体的图像数据集自动采取与标注框架

技术领域

[0001] 本发明涉及机器学习与深度学习领域,特别涉及一种陆地对抗智能体的图像数据集自动采取与标注框架。该框架可以给陆地对抗智能体场景下的目标检测、语义分割、深度估计等各种深度学习任务提供合成的对抗智能体图像数据集。

背景技术

[0002] 神经网络的进步极大地推动了计算机视觉的发展,改变了计算机视觉领域的研究面貌和思维方式。特征工程的思维模式几乎被利用神经网络自动从大量图像中提取特征的方式所取代。这也意味着图像数据成为了计算机视觉领域最重要的资源。计算机视觉领域的一些重要任务如:人脸识别,目标检测,目标追踪,语义分割,场景识别都需要大数据的支撑。所以,图像数据集的数量和质量就成为了计算机视觉关键技术进一步发展并真正走入社会生活和生产的关键。

[0003] 目前,专门针对智能体对抗方面的图像数据集很少,而直接将现有的数据集应用到智能体对抗领域上去会产生数据分布不吻合,深度模型收敛慢,模型泛化能力低以及深度学习算法不鲁棒等各种问题。直接重新收集专门的智能体对抗数据集需要耗费大量的时间,人力和财力,并且还会由于各种特定问题难以进展下去。这些问题限制了人工智能在智能体对抗领域的发展。本发明提出的基于Unity3D的陆地对抗智能体图像数据集自动采取和标注的框架能够非常方便,准确地自动实时收集图片并自动标注图片的深度信息,图片中的物体包围框和图片的语义分割真值,最终得到高质量的智能体对抗合成图像数据集。可以推动人工智能在智能体对抗领域的发展。

发明内容

[0004] 针对现在智能体对抗领域内收集数据集的一些困难,本发明的目的是提供一种陆地对抗智能体的图像数据集自动采取与标注框架,其特征在于,包括:自动实时标注图像深度信息的方法,自动实时标注图像的语义分割标签的方法,自动实时标注图像中物体包围框的方法,图片与图片标签信息的自动实时存储的方法。

[0005] 所述自动实时标注图像深度信息的方法包括以下步骤:

[0006] 步骤S1,从场景中得到各个物体到相机的距离信息d;

[0007] 步骤S2,将距离信息编码为灰度值;

[0008] 步骤S3,Unity3D利用步骤S2得到的灰度值进行渲染得到的场景图片即为深度信息标注图;

[0009] 所述自动实时标注图像的语义分割标签的方法包括以下步骤:

[0010] 步骤S1,给不同类别的物体设置好不同的标签;

[0011] 步骤S2,通过哈希映射的方式,将不同的标签编码为不同的颜色;

[0012] 步骤S3,将物体渲染的颜色参数值设置为步骤S2所得到的颜色值;

[0013] 步骤S4,渲染后的图片就是图像语义分割的标签真值;

[0014] 所述自动实时标注图像中物体包围框的方法包括以下步骤:

[0015] 步骤S1,获取物体在三维场景中的三维包围框和物体的类别标签与类别名称;

[0016] 步骤S2,根据物体间的遮挡关系,利用二分法对步骤S1所得的三维包围框进行裁剪;

[0017] 步骤S3,利用坐标变换关系式将三维包围框的八个顶点投影到相机的二维屏幕上,得到屏幕上的8个点,记为 D_{1-8} ,每个点 D_i 的坐标为 (x_i, y_i) ,根据 D_{1-8} 可以得到二维屏幕上的包围框 $\text{Rect}(X_{\min}, X_{\max}, Y_{\min}, Y_{\max})$,然后根据如下式子得到二维包围框的标注真值;

$$[0018] \quad \begin{cases} C_{id} \\ X_c = X_{\min} + (X_{\max} - X_{\min})/2 \\ Y_c = Y_{\min} + (Y_{\max} - Y_{\min})/2 \\ W = X_{\max} - X_{\min} \\ H = Y_{\max} - Y_{\min} \end{cases}$$

[0019] 其中, C_{id} 为包围框所属物体的类别标签;

[0020] 所述步骤S2进一步包括以下步骤:

[0021] 步骤S21,对于步骤S1所得的三维包围框的每个面,如果面上的四个顶点均被遮挡,就反复执行十次步骤S22,如果面上的四个顶点没有完全被遮挡,则不做处理;

[0022] 步骤S22,记当前所处理的面为 F_i ,与它相对的面为 F_j , F_i 与 F_j 的距离为 d ,判断 F_i 的四个顶点是否都被遮挡,如果是,就将 F_i 朝靠近 F_j 的方向移动 $d/2$,如果否,则将 F_i 朝远离 F_j 的方向移动 $d/2$;

[0023] 步骤S3所述的坐标变换关系式为:

$$[0024] \quad \begin{pmatrix} x_{camera} \\ y_{camera} \\ z_{camera} \\ w_{camera} \end{pmatrix} = \begin{bmatrix} \frac{n}{r} & 0 & 0 & 0 \\ 0 & \frac{n}{t} & 0 & 0 \\ 0 & 0 & \frac{-(f+n)}{f-n} & \frac{-2fn}{f-n} \\ 0 & 0 & -1 & 0 \end{bmatrix} \begin{pmatrix} x_{world} \\ y_{world} \\ z_{world} \\ w_{world} \end{pmatrix}$$

[0025] 其中 n 和 f 分别为相机视锥的近平面距离和远平面距离,而 t 和 r 可以由相机的垂直视角 $fovY$ 及投影屏幕的宽高比 $aspect$ 得出,具体关系由如下式子给出:

$$[0026] \quad \begin{cases} \tan \frac{fovY}{2} = \frac{t}{|n|} \\ aspect = \frac{r}{t} \end{cases}$$

[0027] 所述图片与图片标签自动实时存储的方法包括以下步骤:

[0028] 步骤S1,设置自动图像与标注存储的间隔时间;

[0029] 步骤S2,根据步骤S1的间隔时间,将渲染后的图片,深度信息图片和语义分割标注图片存储到磁盘,并将所标注的物体包围框,物体类别标签和类别名称统一转为Json格式存储到磁盘。

[0030] 由于本发明可以自己创建专门针对陆地对抗的各种虚拟场景,并且可以在虚拟场景中事先设置好物体的类别标签,然后根据类别标签设置好物体的渲染特征,因此在获取图片的时候同时也得到了图片的标签信息。因此本发明提供的框架能够方便,准确而迅速地获取专门针对陆地对抗各种虚拟场景的图片和图片标注信息,从而得到高质量的陆地对

抗智能体合成数据集。

附图说明

[0031] 图1是本发明提出的陆地对抗智能体的图像数据集自动采取与标注框架的结构图。

具体实施方式

[0032] 为使本发明所提的框架的特点更加清楚,框架所用的自动标注方法和自动存储方法的优点更加明显,下面结合附图和具体实施方式做进一步的详细说明。

[0033] 本发明的目的是提出一种陆地对抗智能体的图像数据集自动采取与标注框架,它能够非常方便,准确而迅速地自动收集专门针对陆地对抗各种场景的图像并自动标注图像的深度信息,图像中的物体包围框和图像的语义分割真值,最终得到高质量的陆地对抗智能体合成数据集。

[0034] 图1是本发明提出的基于Unity3D的陆地对抗智能体图像数据集自动采取和标注框架的架构图,该框架可以用在各种专门针对陆地对抗的Unity3D虚拟场景中,下面以一个虚拟的城市坦克智能体对抗场景为例对本发明进行介绍。如图1所示,本发明所提的框架主要包括三大部分:参数初始化部件,图像标注部件和数据存储部件;而图像标注部件又由深度信息标注、语义分割标注、物体包围框标注三个小部件组成,数据存储部件由图像格式存储和文本格式存储两个小部件组成;图中的箭头代表数据流的方向。以下逐个具体地介绍框架的各个部件。

[0035] 参数初始化部件,参数初始化部件的功能是读取配置文件并根据配置文件中的参数设置好框架的基本参数,框架的基本参数包括:数据存储的时间间隔 t_I ,存储图片总数 N ,导出图片的宽度 w ,导出图片的高度 h ,需要标注包围框的物体类别名称列表NameArray(如[soldier,tank])。然后参数初始化部件将 $w,h,NameArray$ 传递给图像标注部件,将 t_I,N 传递给数据存储部件。

[0036] 图像标注部件,接收了从参数初始化部件传来的 $w,h,NameArray$ 后,图像标注部件就开始进行图像的标注。图像标注部件由深度信息标注、语义分割标注、物体包围框标注三个小部件组成,每个小部件都有一个虚拟相机,三个虚拟相机的位置和角度完全一样。下面分别介绍这三个小部件。

[0037] 深度信息标注部件,深度信息标注部件根据自动实时标注图像的深度信息标签的方法对图像进行标注,自动实时标注图像的深度信息标签的方法包括以下步骤:

[0038] 步骤S1,从场景中得到各个物体到相机的距离信息 d ;

[0039] 步骤S2,将距离信息编码为灰度值;

[0040] 步骤S3,Unity3D利用步骤S2得到的灰度值进行渲染得到的场景图片即为深度信息标注图;

[0041] 语义分割标注部件,语义分割标注部件根据自动实时标注图像的语义分割标签的方法对图像进行标注,自动实时标注图像的语义分割标签的方法包括以下步骤:

[0042] 步骤S1,给不同类别的物体设置好不同的标签,这一步在框架运行前搭建虚拟场景的时候完成,搭建虚拟场景的时候给不同类别的物体设置不同的标签。

[0043] 步骤S2,通过哈希映射的方式,将不同的标签映射到不同的颜色。具体地,首先利用Rabin-Karp算法将标签字符串映射为32位的整数H,然后将整数H的第0-7位作为RGBA的B值,将整数H的8-15位作为RGBA的G值,将整数H的16-23位作为RGBA的R值,将整数H的24-31位作为RGBA的A值。

[0044] 步骤S3,对于Unity虚拟场景下的每个物体,将其渲染模式的颜色值设置为该物体的类别标签映射得到的颜色值RGBA。

[0045] 步骤S4,经过Unity渲染完所有的物体后,虚拟相机所显示的图片就是语义分割的标签真值。

[0046] 物体包围框部件,物体包围框部件根据从参数初始化传来的NameArray(在当前虚拟的城市坦克大战的场景中,NameArray=[soldier,tank]),根据自动实时标注图像中物体包围框的方法,将所有类别标签为soldier或者tank的物体在图片中的包围框标注出来,自动实时标注图像中物体包围框的方法包括以下步骤:

[0047] 步骤S1,对每个士兵和坦克,查询Unity后台,得到他们在虚拟场景世界坐标系下的三维包围框BBBox,记三维包围框BBBox;

[0048] 步骤S2,对于每个包围框BBBox,先检查对于相机来说是否可见,如果不可见,就舍弃;如果可见,就根据物体间的遮挡关系,对BBBox进行裁剪得到BBBoxf。

[0049] 该步骤中对BBBox进行裁剪得到BBBoxf的步骤进一步包括:

[0050] 步骤S21,记BBBox的6个面为 F_{1-6} ,对6个面中的每个面都用二分法进行裁剪;以 F_i 为例,记其四个顶点为 V_{1-4} ,依次检查四个顶点是否被其他物体遮挡;具体地,从相机位置沿相机与顶点的连线方向发射一条射线,进行碰撞检测,如果射线在到达顶点之前与其他物体相碰,则说明该顶点被其他物体所遮挡了,否则就说明该顶点未被其他物体遮挡。如果面 F_i 的四个顶点均被其他物体所遮挡,则执行步骤S22;否则就不做任何处理。

[0051] 步骤S22,记目前所处理的面为 F_i ,与它相对的面为 F_j , F_i 与 F_j 的距离为 d ,判断 F_i 的四个顶点是否都被遮挡,如果是,就将 F_i 朝靠近 F_j 的方向移动 $d/2$;如果否,则将 F_i 朝远离 F_j 的方向移动 $d/2$;如此反复执行10次。

[0052] 步骤S3,利用(3)式将裁剪后的三维包围框BBBoxf投影到相机的二维屏幕上。具体地,记BBBoxf的八个顶点为 A_{1-8} ,利用Unity3D的坐标变换,分别将 A_{1-8} 投影相机的二维屏幕上,得到二维屏幕上的8个点,记为 D_{1-8} ,每个点 D_i 的坐标为 (x_i, y_i) ;根据以下式子:

$$[0053] \quad \begin{cases} X_{min} = \min_{1 \leq i \leq 8} x_i \\ X_{max} = \max_{1 \leq i \leq 8} x_i \\ Y_{min} = \min_{1 \leq i \leq 8} y_i \\ Y_{max} = \max_{1 \leq i \leq 8} y_i \end{cases} \quad (1)$$

[0054] 可以得到二维屏幕上的包围框Rect($X_{min}, X_{max}, Y_{min}, Y_{max}$)。然后根据式(2)得到二维包围框的标注真值

$$[0055] \quad \begin{cases} X_c = X_{min} + (X_{max} - X_{min})/2 \\ Y_c = Y_{min} + (Y_{max} - Y_{min})/2 \\ W = X_{max} - X_{min} \\ H = Y_{max} - Y_{min} \end{cases} \quad (2)$$

[0056] 其中 C_{id} 为包围框所属物体的类别标签。

$$[0057] \begin{pmatrix} x_{camera} \\ y_{camera} \\ z_{camera} \\ w_{camera} \end{pmatrix} = \begin{bmatrix} \frac{n}{r} & 0 & 0 & 0 \\ 0 & \frac{n}{t} & 0 & 0 \\ 0 & 0 & \frac{-(f+n)}{f-n} & \frac{-2fn}{f-n} \\ 0 & 0 & -1 & 0 \end{bmatrix} \begin{pmatrix} x_{world} \\ y_{world} \\ z_{world} \\ w_{world} \end{pmatrix} \quad (3)$$

[0058] 其中 n 和 f 分别为相机视锥的近平面距离和远平面距离,而 t 和 r 可以由相机的垂直视角 $fovY$ 及投影屏幕的宽高比 $aspect$ 得出,具体关系由如下式子给出:

$$[0059] \begin{cases} \tan \frac{fovY}{2} = \frac{t}{|n|} \\ aspect = \frac{r}{t} \end{cases} \quad (4)$$

[0060] 数据存储部件,数据存储部件从参数初始化部件获得数据存储时间间隔参数 t_1 和所需存储图片总数 N ,并根据 t_1 来确定当前时间是否需要将图像标注部件的标注信息存储到磁盘。如果当前时间需要存储,数据存储部件则将原图片,图片的深度信息标注,图片的语义分割标注以图片格式存储到磁盘。将图片的物体包围框标注真值(C_{id}, X_c, Y_c, W, H)以Json文本格式存储到磁盘。然后数据存储部件判断当前已经存储的图片数目是否超过所需存储图片总数 N ,如果超过,则框架停止运行,如果没有超过,则框架继续运行。

[0061] 最后我们就得到了虚拟城市坦克智能体对抗场景的图像数据集,包含RGBA图片,图像深度信息标签,图像语义分割标签和图像中物体包围框标签。

[0062] 以上所述的具体实施方法,对本发明的目的,技术方案和有益效果进行了详细说明,所应理解的是,以上所述仅为本发明的具体实施例而已,并不用于限制本发明,凡在本发明的精神原则之内,所做的任何修改、等同替换,改进等,均应包含在本发明的保护范围之内。

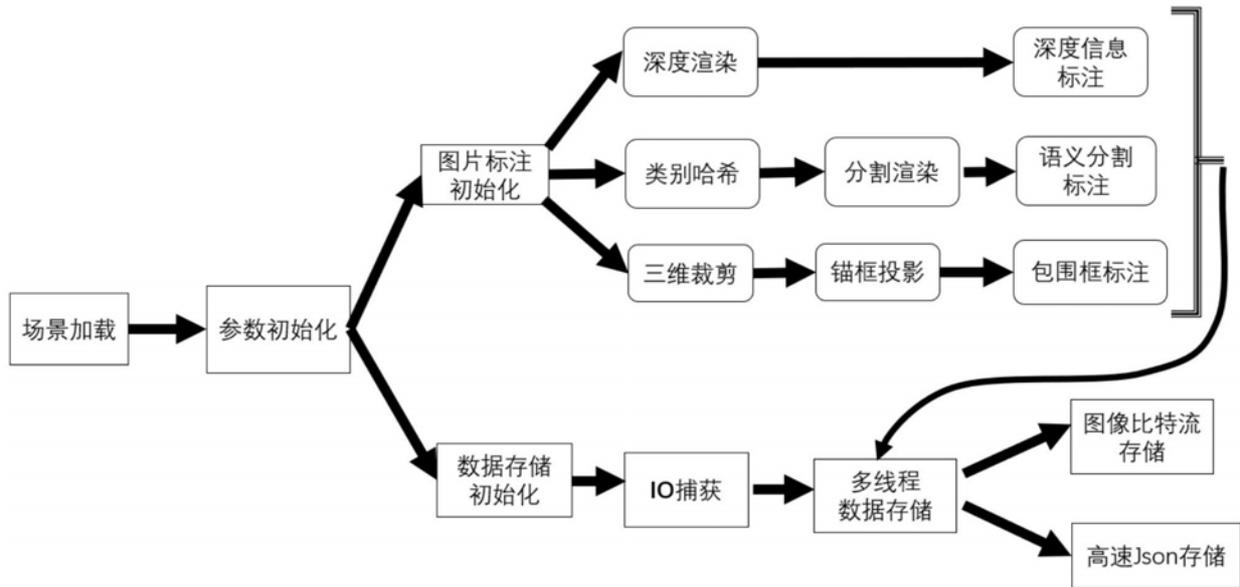


图1