arXiv:2304.04901v2 [cs.CV] 6 Nov 2024

通过在线视觉反馈高效收集二维目标检测训练数据集

Takuya Kiyokawa*、Naoki Shirakura、Hiroki Katayama、Keita Tomochika、Jun Takamatsu

摘要——训练基于深度学习的视觉系统需要对大量图像进行人工标注。这种人工标注既耗时又费力。尽管先前的研究试图消除标注所需的工作，但图像收集所需的工作仍然存在。为此，我们提出一种基于 Web 应用程序的人在回路数据集收集方法。为了平衡工作负荷与绩效，我们通过鼓励用户以有趣的方式收集多视角物体图像数据集来提升其积极性，并提出三种在线视觉反馈功能，以跟踪收集状态的进展。我们的实验全面考察了各项功能对采集性能和操作质量的影响。结果表明，标注和目标检测具有可行性。

I. 引言

近年来，采用深度学习（DL）的高精度目标检测系统在物流 [1], 废物处理 [2], 制造业 [3], 和零售业 [4]等领域受到广泛关注。这些行业中新目标物体会接连快速出现，因此需要及时收集新的训练数据集，并重新训练目标检测模型。然而，由于收集训练数据集需要耗费大量时间和精力，从引入新物体到执行目标检测的整个流程极难完成。这一问题不仅对仅使用相机的应用（如产品检测、垃圾分选机、机械零部件缺陷检测、自动收银机和家用机器人）十分紧迫，对于其所在领域中基于视觉的机器人系统的自动化操作同样如此 [5]–[10]。

在收集训练数据集时，通常需要使用标注工具，手动在图像中的物体周围绘制二维（2D）边界框，并仔细标注每个二维边界框所包含的物体。这一标注过程既耗时又费力。带噪声的标注会降低训练所得模型的绩效 [11]。由于现有数据集的质量较低（例如一致性较差），对其进行人工标注一直被认为存在问题 [12],，也有人尝试通过更加费力、遵循指南的重新标注来解决这一问题 [13]。人工标注过程存在一定的主观性（尤其是在存在遮挡的情况下），而且标注人员感到疲劳时，标注质量可能会下降。

众包是一种高效且经济的图像数据集标注采集方法。为控制众包标注的质量，人们提出了人在回路中的验证技术 [14],基于 Web 的图像标注界面 [15],[16], 以及基于 Web 的视频标注平台 [17], [18]。然而，人为错误始终难以避免，因此训练出的深度学习模型未能达到最佳性能。

为了解决这一问题，Kiyokawa 等人 提出了一种利用视觉标记作为工具、从图像中自动获取标注的方法 [5], [7]。他们将标记贴在每个物体上，并在现实世界中将贴有标记的物体摆放成不同姿态。由于视觉标记很容易检测，该方法能够自动标注给定的物体与标记之间的位置关系。在自动标注过程中，标记用于目标检测。然而，如果将显示标记的图像用于训练，神经网络通常会将标记学习为物体的一项特征。通过用噪声掩膜遮挡标记，我们成功减少了这种错误学习。

然而，即使采用这种方法，也有必要快速收集从各个方向拍摄的图像。我们通过利用在线视觉反馈，让不同的人执行这项任务，从而解决这一问题。基于视觉标记的自动标注，即使存在遮挡，也能保持标注的一致性。除了标注流程之外，Kiyokawa等人 还开发了一种可自动改变物体姿态的系统 [19]。利用手眼机器人和旋转台，固定在机械臂末端的相机可以通过移动相机和物体，拍摄处于各种姿态的物体。

本研究旨在通过提升个体用户的愉悦动机，提高大规模图像数据集收集的整体能力，而不是单纯投入大量人力资源。如果能够借助任何用户都能使用的应用程序，而非以往方法 [19], 中使用的特殊设备，高效地收集图像数据集，那么数据集收集方法将迎来范式转变。由于在线视觉反馈是通过我们开发的增强现实标记实现的自动标注，因此在 Web 应用程序中很容易实现。

本研究旨在开发一种紧凑收集多视角物体图像数据集的方法，这是一款 Web应用程序，能够让用户轻松捕捉自由视点的图像。

T. Kiyokawa 任职于日本大阪府丰中市大阪大学工程科学研究科系统创新系。

*通讯作者：kiyokawa@sys.es.osaka-u.ac.jpN. Shirakura 任职于日本东京江东区青海的国立产业技术综合研究所。H. Katayama、K. Tomochika 和 J. Takamatsu 任职于日本奈良县生驹市奈良先端科学技术大学院大学（NAIST）信息科学部。

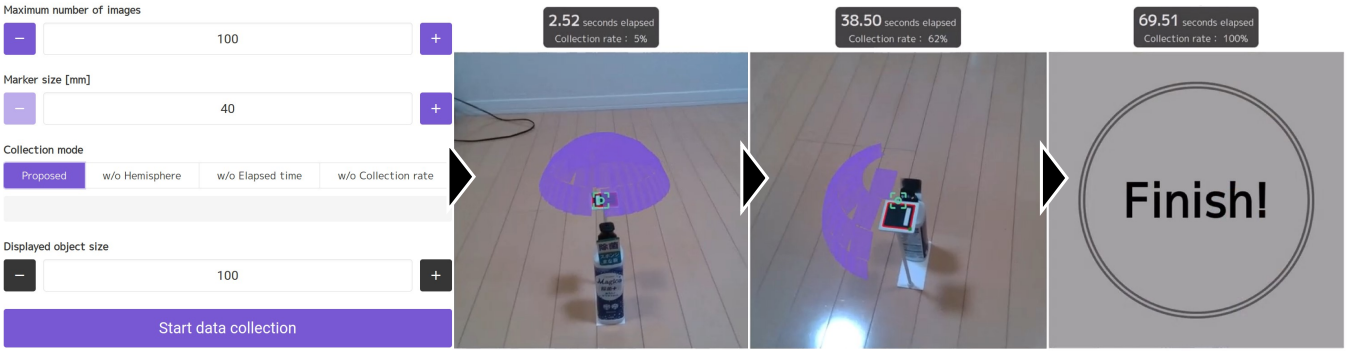


图 1：所开发的移动应用程序的显示屏切换。

现实世界中的图像。用户可以通过在目标物体周围移动相机设备来采集目标物体的图像，之后系统会自动进行标注。图 1展示了应用程序显示屏的一个示例，其中显示了提升至100%的采集率（定义为用户指定的目标数量）、带有半球形物体的未采集视点，以及从采集开始起的已用时间。

为了鼓励用户参与人在回路数据集收集，提升采集性能和操作质量至关重要。本文所称的操作质量，是指单个操作人员对收集操作标准的感知，涵盖收集操作的情感、社交和身体层面，包括工作负荷、享受度和时间压力。我们实现了三个简单而有效的应用程序功能，用于展示数据集采集状态。以下各项是所提出的应用程序功能，表 I 汇总了这些提案对采集性能和操作质量的预期效果。

Hm半球：将未采集区域表示为半球形状，可以帮助用户找到相机应移动的方向，从而以愉悦的方式采集无偏的图像数据集。 **CR采集率：** 显示采集率可以向用户反馈采集是否成功，从而激发用户对任务的愉悦动机。 **ET已用时间：** 显示已用时间可以鼓励用户在时间压力下高效采集数据集，尽管这可能会给用户带来一些心理负担。此外，所提出的Web应用程序还汇总并显示了最终已用时间的排名，从而使采集工作更具游戏化效果。

ALL **全功能**：模式包含上述三个功能。

对于基于深度学习的目标检测器数据集，有必要从多个视点采集图像，以增加数据集的多样性。将两类位置（已采集和待采集）以半球的形状显示出来，并将采集过程游戏化，让用户快速擦除半球的所有表面，就能同时提升数据集的质量。此外，通过

表 I：应用功能的预期影响。

	采集性能		操作质量		
	效率	多样性	工作负荷	愉悦感	时间压力
Hm	-	✓	-	✓	-
CR	-	-	-	✓	-
ET	✓	-	✗	✓	✓

在工作过程中提供采集率，并在应用程序执行结束时提供收集时间排名功能，用户可能会将该应用程序视为一款游戏，目标是快速擦除半球。

本研究考察了所提出的三项功能对用户心理工作负荷、可用性、愉悦动机、效率和数据集质量的影响。该分析涵盖了如表所示的潜在积极和消极后果。在实验中，为了衡量每项功能的效果（例如，Hm、CR 和 ET），我们通过与四种方法进行比较开展了消融研究。其中一种方法包含全部三项功能（ALL），另外三种方法则分别移除其中一项功能。

首先，七位参与者在使用应用程序后立即填写了有关心理工作负荷的 *NASA*任务负荷指数（NASA-TLX）[20] 问卷、有关可用性的系统可用性量表（SUS） [21] 问卷，以及一份关于个人信息、先前知识、以往经历和令人愉悦的功能的自定义问卷。其次，为了评估效率以及收集无偏数据集的可能性，我们测量了数据集收集试验所用的时间，并计算了所收集数据集的多样性，以评估效率和数据集偏差。本研究从两个方面为相关文献作出了贡献：

1) 为了扩大自动标注物体图像的适用范围，我们提出、开发并评估了一款用于采集多视角图像数据集的移动应用程序。使用所开发的系统，本研究阐明了标注和目标检测的可行性（**第 6 节**）。2) 通过全面评估，我们阐明了融入所提议功能（Hm、CR 和 ET）的可能性，可以

以令人愉快的方式收集数据集，从而降低心理工作负荷和时间压力，并提高数据集的多样性（**第 5 节和第 7 节**）。

II. 相关研究

本节介绍了有关利用在线视觉反馈技术有效收集训练数据集，以及基于云的数据集收集优势的相关研究。

A. 数据集收集中的在线视觉反馈

在计算机视觉领域，为了给网络图像标注恰当的标签，以实现更准确的图像搜索，Ahn 等人 [22] 推出了一种交互式系统：一款有趣的游戏，可用于为网络图像提供有意义的标签。Deng 等人 [23]提出了一款名为 *bubbles* 的游戏，它揭示了人类使用的具有判别力的图像特征。玩家的目标是识别严重模糊图像的类别。他们还提出了一种图像识别算法*Bubble Bank*，该算法利用人类选择的气泡来提升识别绩效。

与我们的目标类似，Kavasidis 等人 [24] 推出了一款闪游，旨在轻松生成用于测试鱼类检测器的真实标注。他们开发了一款名为*Flash theFish* 的在线游戏，用户会看到水下环境的视频，并须通过点击鱼来为其拍照。该方法依赖于手动点击图像中的鱼。将所得结果与手绘真实标注进行比较后发现，一款能够隐式控制用户质量的游戏可以轻松创建可靠的图像和视频标注。然而，这款游戏仅限于鱼类检测，而且游戏开发者必须投入大量精力，为每种特定领域的检测模型设计游戏场景。

在另一项针对目标检测数据集创建的研究中，Fusiello 等 [25] 提出了一种检测并标注图像中物体的方法。他们描述了一款名为 *Ask'nSeek* 的双人网页版猜谜游戏，以一种有趣且互动的方式支持这些任务。Ask' nSeek 要求用户根据语义和拓扑线索猜测图像中隐藏区域的位置；因此，这是一种不同于我们方法的半自动标注方法。游戏日志中收集的信息与图像内容分析算法的结果相结合，并用于训练机器学习算法，该算法输出图像中最相关区域的轮廓及其名称。

他们阐明了用于图像标注的游戏的可行性-
标注。然而，如前所述，这些方法 [24]‚[25] 无法自动标注真实世界图像。

B. 基于云的数据共享

一些研究人员研究了云机器人技术的潜在优势，例如大数据、云计算、集体机器人学习和人类计算，相关讨论和综述见 [26]–[28]。

在机器人操作领域，利用云计算和

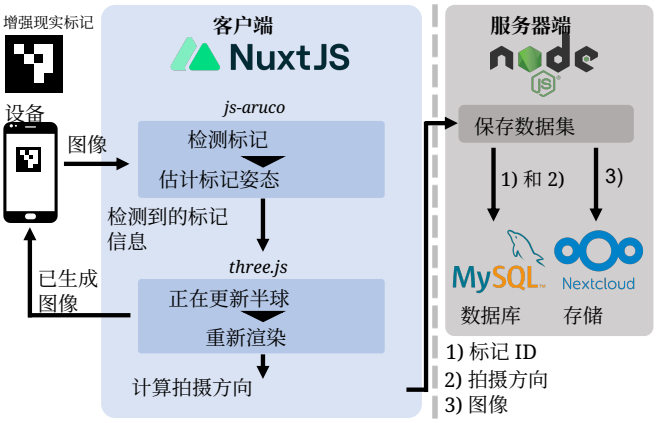
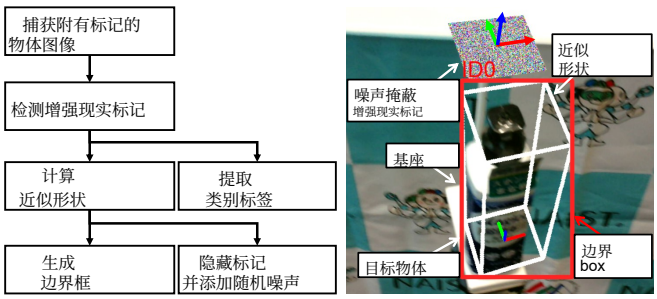


图 2：移动应用的系统配置

ion.



(a) 标注流程　　(b) 叠加的标注

图 3：自动二维边界框标注流程。

互联网服务的抓取和视觉系统迄今已被提出。以往研究证明了实际云机器人应用的可行性。一个尚待解决的重要问题是，如何构建一种有效的数据集收集方法，用于训练尤其适用于云机器人框架的系统。如果没有一种能够轻松、快速上传至互联网的数据集收集方法，我们就无法假设互联网上已有大量数据库；因此，当各种新物体出现时，要及时作出响应将十分困难。在这些数据集收集问题中，本研究以基于云的机器人视觉系统为前提，聚焦于物体图像数据集收集。

III. 人在回路数据集收集

A. 系统实现

图 2 展示了移动应用程序的系统配置。我们实现了一个 Web应用程序，其中包含一个客户端组件，用于收集和标注图像、可视化收集状态，以及一个用于存储生成的数据集的服务器端组件。一般来说，Web应用程序不受设备类型限制，包括 Windows 平板电脑和 Android 智能手机等移动设备。这些移动设备通常配备额外的相机，安装在设备上的 Web应用程序可通过移动设备的相机，从自由视点轻松采集多视角图像。设备的计算资源用于执行图像采集、标记。

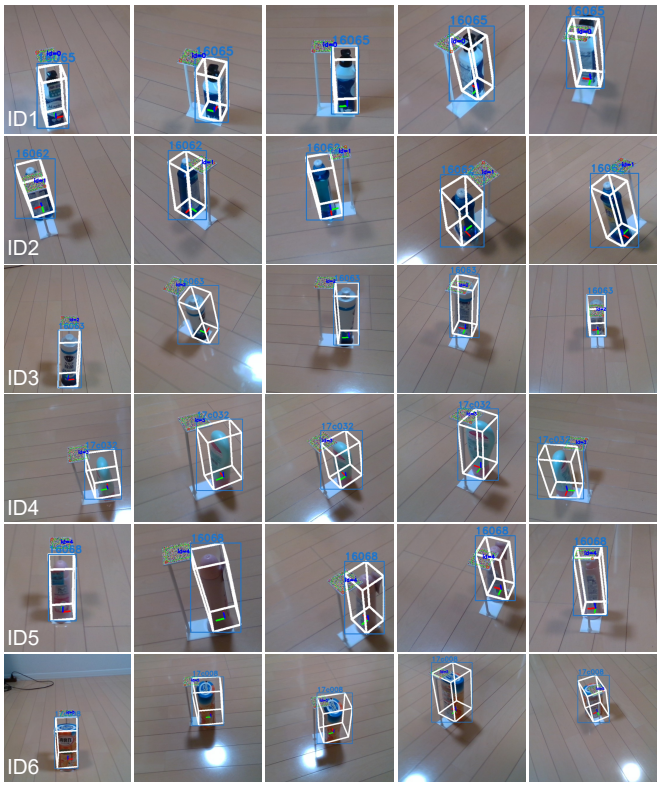


图 4：收集的数据集的可视化。

检测、标记位姿估计以及显示屏上的各种绘图。

为了自动标注物体图像中的二维边界框，我们采用了一种基于增强现实标记的标注方法 [5]。图 3（**a**）展示了流程。在启动应用程序之前，用户将目标物体固定在标记夹具上，并将夹具放置在相机能够从物体周围拍摄到该物体的位置。参与者使用经过预先校准的相机。校准使用 NxM 棋盘格和 OpenCV 的 calibrateCamera 方法完成。由于假定物体与夹具上的标记之间具有已知的相对位姿，且与物体类别关联的标记编号（ID）已知，因此可以在 Web 应用程序运行期间自动完成标注。图 3（**b**）展示了收集的数据集的可视化，而图 4 则展示了从不同视点拍摄的其他物体示例。

我们使用*Nuxt.js*[1] 以及多个 JavaScript 库来实现各种功能，包括增强现实（AR）库*js-aruco*[2],和 3D 库 *three.js*[3]。js-aruco 可检测捕获图像中的增强现实标记，并估计标记的姿态。three.js是一个用户友好的库，可以在网页上渲染 3D内容，进而通过更新姿态，为更新捕获图像中半球的外观提供支持，移除矩形上的

[1]https://nuxtjs.org/ [2] https://github.com/jcmellado/js-aruco [3] https://threejs.org/

刚刚捕获的半球，并应用它再次进行渲染。最后，渲染后的图像显示在移动设备的显示屏上。在客户端的一次循环中，服务器将图像数据集保存到云存储中，并通过 *MySQL*[4] 数据库和*NextCloud*[5] 在线存储服务进行结构化管理。

图 1展示了显示屏切换的示例。在初始显示屏上，用户设置四个参数：目标图像数量、附着在物体上的视觉标记的大小、要执行的采集模式，以及要显示的半球形物体的大小。设置这些参数后，用户即可开始采集过程。所选模式一启动，用户便可以通过显示所提出的状态指示器来查看捕获的图像，包括以半球（Hm）表示的采集进度、采集率（CR）和已用时间（ET）。达到用户指定的图像数量后，图像捕获将自动结束。

B. 通过状态可视化提供在线视觉反馈

起始点处的显示屏会显示所选模式的说明以及激活该模式所需的参数。点击图1 最左侧图像底部标记为 开始数据采集的按钮后，倒计时开始，直至图像捕获启动。从五秒倒数至零后，图像捕获和标记检测同时开始。在图像捕获期间，显示屏上会出现特定于模式的可视化结果，其中包括显示采集进度的ET、CR 和 Hm。当采集率达到 100% 时，图像捕获和标记检测结束，并且显示屏上会显示消息完成！ 。图像捕获完成后，用户可以在Web应用程序的顶部窗口中查看收集时间排名。

以下各节将详细介绍所提出的三项应用程序功能（表 I）：显示已用时间、显示采集率，以及将未采集区域表示为半球。

1）以半球形式绘制采集进度：为在采集试验过程中隐式地向用户显示采集进度，应用程序窗口会显示一个带有缺失部分的半球，以表示尚未捕获的区域。半球被划分为一定数量的矩形（数量可通过图 1 最左侧图像中所示的显示物体大小进行设置），并首先显示划分后的矩形物体。为使矩形彼此不重叠，它们沿半球表面呈螺旋状排列，如图 1 所示。半球上的矩形按照 [32] 所述进行排列。

使用*setFromSphericalCoords Vector3* 类[6] 中的three.js，我们可以轻松计算物体在笛卡尔坐标系中的位置 $\boldsymbol{v}$ ，如下所示。

$$\boldsymbol{v} = (r\sin\phi\sin\theta, r\cos\phi, r\sin\phi\cos\phi)\,, \tag{1}$$

[4]https://www.mysql.com/ [5] https://nextcloud.com/ [6]https://threejs.org/docs/#api/en/math/Vector3

其中 r 表示半球半径。为了使矩形物体的平面法向量与 v平行，我们使用 three.js 的 *lookAt* 函数，该函数属于*Object3D* 类[7] 。

当所捕获图像的中心出现一个矩形物体时，相应的显示矩形便会消失，表示采集已完成。半球会尽可能少地遮挡物体的外观。在渲染显示的图像上叠加一个半透明的半球形物体。这些半透明物体使用户能够轻松观察实际目标物体的姿态。用户不仅可以通过采集率监控采集进度，还可以通过观察不断更新的半球来了解进度；该半球提供有关尚未捕获区域的信息，并帮助用户有效移动相机。借助此功能，视点（即：从相机观察到的物体姿态）不会与现有数据集重叠，从而生成无偏数据集。

显示半球可以在采集过程中向用户提供反馈。这会带来类似游戏的采集感。用户可以一边享受采集图像的过程，一边思考如何尽快移除半球。这类游戏化功能会激励用户快速采集数据集。

2）显示数值采集百分比：图 1 展示了执行过程中显示的采集率。计算得到的采集率显示在显示屏顶部、图像捕获窗口上方。显示采集率同样会带来类似游戏的采集感。用户可以一边享受采集图像的过程，一边思考如何提高采集率。

采集率是根据先前收集的数据集中包含的六维（6D）物体位姿数据（相对于相机视角）计算得出的。实际上，当半球表面上显示的矩形位于相机拍摄图像的中心时，目标矩形便会消失，获取的图像数据集也会保存。因此，在采集 100 张图像时，采集率会增加 1%。这种方法可以减少从相机视角看数据集中的重复内容。通过让用户自行避免数据集重复，这项功能带来了舒适感和可玩性。

半球图能够更详细、直观地展示这一进度。与半球图相比，采集率可以定量衡量目前已收集的数据相对于最终目标值（100%）的完成程度。

3) 提供已用时间和排名： 在模式执行期间显示的已用时间会促使用户在时间压力下更快地收集数据集，尽管这可能会给用户带来一些精神负担。此外，所提出的Web应用程序汇总并显示最终已用时间的排名，以激励用户像玩游戏一样愉快且高效地进行收集。收集时间越短，排名就越高。

[7]https://threejs.org/docs/#api/en/core/Object3D

在排名表中，我们可以查看此前试验在所有试验中的排名、模式名称、绩效（采集时间 [s] /图像数量）、采集时间 [s], 以及图像数量。

IV. 用户体验评估

A. 实验方案

我们的用户实验评估了使用所提出的应用程序进行数据集采集试验时的心理工作负荷、可用性、愉悦动机、效率（收集时间）和数据集质量（多样性）。一个应用程序（ALL）集成了三项功能（即Hm、CR 和 ET），并以单一应用模式实现。我们还开发了另外三种应用模式，每种模式分别缺少这三项功能中的一项：一种不含 Hm，一种不含 CR，另一种不含 ET。所有已实现模式下，为目标物体收集数据集的典型流程如下：

1) 在目标物体附近易于检测到该物体的位置贴上视觉标记 2) 在小型手持笔记本电脑（微软 Surface PRO）上启动 Web应用程序 3) 从四种模式中选择一种（我们事先为每位参与者随机生成了模式顺序） 4) 启动该模式，在采集多视角图像的同时查看显示屏上的可视化结果 5) 之后该模式结束（如果采集率达到 100%），参与者将排名与其他采集结果一同查看 6) 参与者填写包括 NASA-TLX、系统可用性量表（SUS）和自定义问卷在内的问卷 7) 对四种模式中的每一种，重复 3) 至 6) 的流程三次

我们调查了来自日本奈良先端科学技术大学院大学的七名成年男性。参与者填写了问卷，如表II所示。由于新冠疫情的防控措施，我们只能邀请来自同一所大学的参与者参加实验。不过，我们尽可能从不同背景中招募受试者开展实验。从表 III 所示的问卷回答可以看出，我们招募了具有不同专业领域的参与者。表 IV 显示了每位参与者的实验顺序。

启动该模式后，采集到 100 张图像时，采集率达到 100%，执行的模式随即结束。如图 1 所示，附着在物体上的视觉标记尺寸和要显示的半球形物体尺寸分别设为 40 毫米（边长）和 100 毫米（半径）。

B. 评估指标

第5节 从多个方面讨论这些结果。首先，我们分析了用于评估心理工作负荷（**第5节.1**）和可用性（**第5节.2**）的问卷结果，以了解主观视角。参考 [33],, 我们询问参与者在试验期间的使用情况，以比较各项功能的效果。

表II：与 NASA任务负荷指数和系统可用性量表分开设计的定制问卷。

问题	Type	选项
性别	多选题	男性或女性
Age	开放式	正整数
您体验过 AR 技术吗?	已关闭	否或是
你对训练数据集收集了解多少?	多项选择	什么都没有，一点点，嗯，或者你的专业
你觉得它像游戏一样有趣吗？你觉得哪些功能很有趣?	多项选择题	不喜欢它、Hm、CR、ET 或排名
你是否努力收集了数据集？哪些功能激励了你？多项选择题		不这么认为，嗯，CR、ET 或排名

表 III：关于参与者背景的回答。

	参与者 ID						
问题 ID	1	2	3	4	5	6	7
1†a	No	No	Yes	Yes	Yes	Yes	Yes
2†b	4	2	2	3	2	4	4

†a 您体验过 AR 技术吗? †b您对训练数据集收集了解多少? 回答数字 1、2、3 和 4 分别对应完全不了解、略知一二、比较了解和专业领域。

表 IV：每位参与者的实验顺序。

	实验顺序			
参与者 ID	1	2	3	4
1	不含 ET×3	不含 CR×3	ALL×3	不含 Hm×3
2	无 Hm×3	无 CR×3	ALL×3	不含 ET×3
3	无 Hm×3	无 ET×3	无 CR×3	ALL×3
4	不含 CR×3	不含 ET×3	不含 Hm×3	ALL×3
5	ALL×3	不使用 Hm×3	不使用 ET×3	不使用 CR×3
6	不含 ET×3	ALL×3	无 CR×3	无 Hm×3
7	不含 CR×3	ALL×3	不使用 Hm×3	不使用 ET×3

在每种模式中均已实现。为评估心理工作负荷和可用性，七名参与者分别填写了NASA任务负荷指数和系统可用性量表。为阐明所实现功能对用户源于愉悦感的动机的贡献，参与者回答了表II（**第5.3节**）中列出的自定义问卷。其次，我们从收集时间（**第5.4节**）以及所收集数据集的多样性（**第5.5节**）两个方面评估了所提出的Web应用程序的效率。

V. 结果

A. 心理工作负荷

图5展示了NASA-TLX问卷的结果。TLX评分以0至100之间的数值表示用户的心理工作负荷，数值越大，工作负荷越高。数值越低，绩效越好。图6和图7展示了NASA任务负荷指数两个不同子量表的评分，包括时间需求和绩效。

红色“x”标记表示平均值，误差条的上限和下限分别表示最大值和最小值。上方

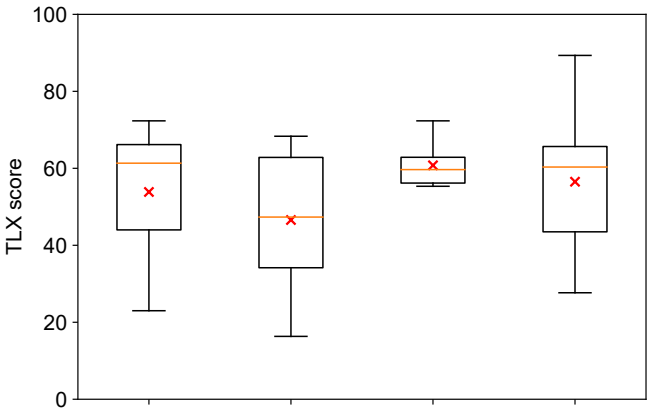


图5：TLX评分。

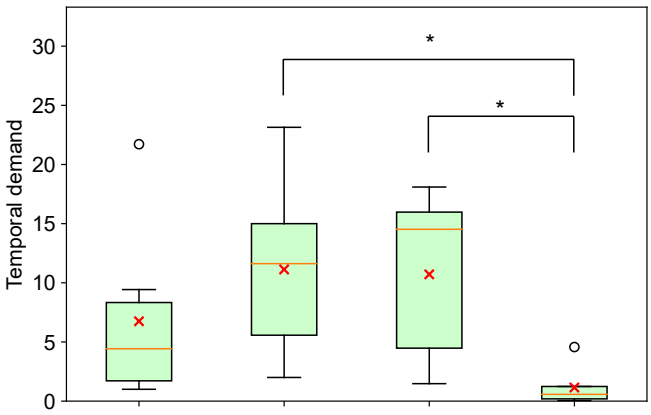


图6：“时间需求”（NASA任务负荷指数的一个子量表）。

箱体边缘表示第75百分位数，水平线表示中位数，底部边缘表示第25百分位数。大于第75百分位数+1.5× 四分位距（IQR ＝75百分位数 − 25百分位数）或小于第25百分位数 −1.5× IQR 的样本被视为离群值（白色圆圈）。

此外，我们进行了多重比较检验，以比较问卷结果。表V展示了 NASA任务负荷指数各子量表的多重比较结果。为进行多重比较，采用了*Dunn-Bonferroni*检验 [34]

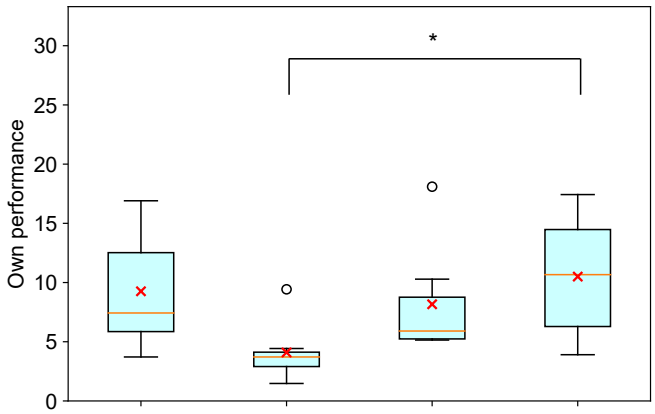


图 7：“绩效”（NASA TLX 的一个子量表）。

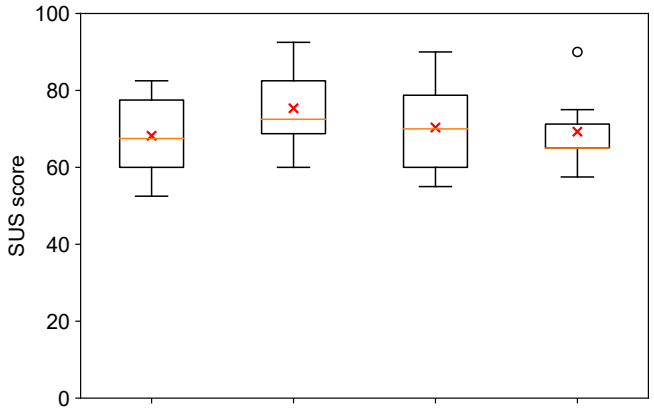


图 8：系统可用性量表得分。

用于统计软件*SPSS*[8]。表V中显示的p值是经Bonferroni校正后的值，上限为1.0。带星号的p值表示方法1和2之间存在显著差异，显著性水平为0.05，这一点也显示在图6和图7中

我们发现 NASA任务负荷指数各子量表的得分存在若干差异，尤其是时间需求 和 表现。不含 Hm 的方法和不含 CR 的方法在时间需求子量表上的评分高于不含 ET 的方法，p 值分别为 0.010 和 0.010，这表明在这两种方法中，用户感受到的时间需求较高。ALL 方法与不含 ET 的方法之间未发现显著差异，p 值为 0.14。然而，七名受试者报告的 ALL 方法时间需求得分均高于不含 ET 的方法。这一结果表明，采用 ET 的方法中，与心理工作负荷相关的某个因素相对较高。

尽管无法确认 ALL 方法与不使用 Hm 和

[8]https://www.ibm.com/spss

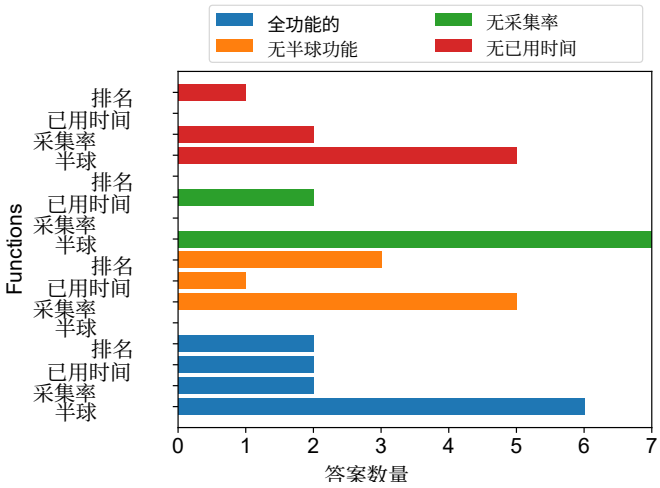


图 9：对自定义问题“你是否像玩游戏一样享受其中？”和“你觉得哪些功能令人愉快？”的回答数量。

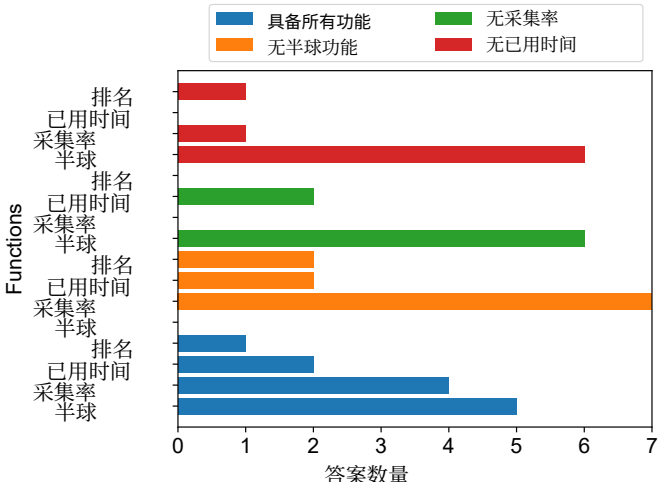


图 10：对自定义问题“你是否努力收集数据集？”和“哪些功能激励了你？”的回答数量。

CR 的方法之间存在显著差异，但可以确认，有五位参与者认为，不使用 Hm 和不使用 CR 的方法比 ALL 方法有更高的时间需求。这一结果表明，在 ALL 方法中结合 Hm 和 CR 等在线视觉反馈功能，可能有助于降低时间需求。这两项功能都用于显示收集操作的进度。正如 Myers [35] 所展示的，用户通常更喜欢进度指示器，因为它们能让新手确信任务已被接受并正在顺利进行，同时为专家用户提供足够的信息，以预测任务的大致完成时间。在基于移动应用程序的数据集收集中，进度指示器带来的安心感也可能有助于降低时间需求。

未采用 Hm 的方法在子量表 绩效 上的得分显著低于未采用 ET 的方法，p 值为 0.030。子量表 绩效 反映用户对自身收集试验的自我评价程度。

得分越高，表示自我评价越差。因此，结果表明，与未采用 ET 的方法相比，未采用 Hm 时对绩效的自我评价往往更好。这种自我评价的改善可能是由于半球实时擦除所提供的收集结果的明确视觉反馈。

B. 可用性

图 8 展示了系统可用性量表问卷的结果。系统可用性量表得分以 0 到 100 的数值表示用户对可用性的评价，分数越高表示可用性越高。图 8 的图表格式与其他图（图5 - 图7）相同。表V还展示了系统可用性量表得分的多重比较结果。各方法之间没有显著差异。

C. 愉悦动机

图 9 和图 10 展示了关于愉悦动机的自定义问卷结果。图9显示了对自定义问题“你是否像玩游戏一样享受其中？”和“你觉得哪些功能令人愉悦？”的回答数量。图 10 显示了对自定义问题“ 你是否努力采集数据集？ ”和“哪些功能激励了你？”的回答数量。选择“更新 Hm”的人数最多，因为它能带来愉悦感，也容易激发动力。在两类问题中，向用户通知采集率的方法在四种方法中排名第二。表 II 列出的其他自定义问题结果并不突出，但排名和已用时间也让用户感到有趣并受到激励。

正如预期，这些结果表明，显示正在更新的 Hm和 CR 可能有助于通过愉悦动机促进数据集收集中的在线视觉反馈。

D. 收集时间

表VI显示了七名参与者在三次试验中每次收集100幅图像所用时间的均值和标准差。此外，我们通过将最终试验的收集时间除以第一次试验的收集时间，计算出了增减率（ID率）。表VI还显示了所有参与者ID率的均值和标准差。该数值与用户对收集操作的熟悉程度以及随着试验次数增加而累积的工作负荷有关。负值表示存在熟悉效应，而正值表示存在疲劳。

在 ALL 模式下，耗时不到 10 [分钟]在三次试验中收集数据，这表明收集 300 张图像（每次试验收集 100 张图像）只需 10[min] 。与其他方法相比，ALL 方法的收集时间均值长于不使用 Hm 的方法，短于不使用 ET 的方法，并且与不使用 CR 的方法几乎相同。对于使用 Hm 的方法，由于用户倾向于仔细检查目标矩形是否消失

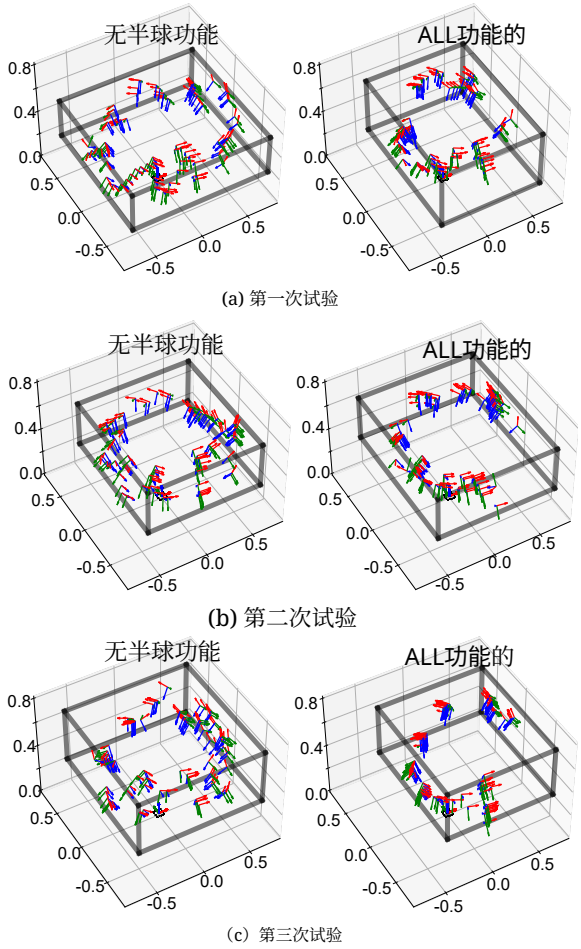


图 11：收集的数据集的多样性。 $z = 0$ 处红色、蓝色和绿色箭头所示的坐标系表示标记的坐标系。其余坐标系表示拍摄图像时的相机位姿。灰色边界框表示包含所有相机视点的最小长方体。

在模式执行过程中，因此与不使用 Hm 的方法相比，所需时间相对较长。

在不使用 Hm 的方法中，随着相机视角位姿的变化，采集率得到改善；因此，采集率更容易提高。同时，除不使用 Hm 的方法外，其余三种方法的收集时间和 ID率几乎相同。各次试验中，最短收集时间的最大差异为 37.5%（72 [s]）。在不使用 ET 的方法中，与 ALL 方法对应的时间相比，第 1 次和第 3 次试验的收集时间更长，且 ID率略高。ID率的结果与收集时间的结果呈现相同趋势。

E. 多样性

为了减少目标物体周围未被采集的区域，收集多视角图像数据集非常重要。为了验证 Hm 对无偏数据集收集的影响，我们比较了 ALL 方法生成的数据集与未使用 Hm 的方法生成的数据集。请注意，在该过程中。

表V：多重比较检验。

方法1	方法2	p值							
		NASA任务负荷指数	MD[†a]	PD[†b]	TD[†c]	OP[†d]	EF[†e]	FR[†f]	SUS
ALL	不含CR	1.0	1.0	0.72	1.0	1.0	1.0	1.0	1.0
ALL	无 ET	1.0	1.0	1.0	0.14	1.0	1.0	1.0	1.0
ALL	无 Hm	1.0	1.0	1.0	1.0	0.14	1.0	1.0	0.38
不含 CR	不含 ET	1.0	1.0	0.47	0.010*	1.0	1.0	1.0	1.0
不使用 CR	不含 Hm	1.0	1.0	1.0	1.0	0.30	0.10	1.0	1.0
不使用 ET	不使用 Hm	1.0	1.0	1.0	0.010*	0.030*	0.47	0.38	0.72

†a 是“心理需求”的缩写，它是 NASA TLX 的一个子量表。†b是“身体需求”的缩写，它是 NASA TLX 的一个子量表。†c 是“时间需求”的缩写，它是 NASA TLX 的一个子量表。†d是“绩效”的缩写，它是 NASA TLX 的一个子量表。†e 是“努力程度”的缩写，它是 NASA TLX 的一个子量表。†f 是“挫折感”的缩写，它是 NASA TLX 的一个子量表。* 星号标注的 p 值表示两种方法之间存在显著差异，显著性水平为 0.05。

表 VI：七位参与者在三次试验中分别收集 100 张图像所用的收集时间 [s] 。每个元素显示时间的均值 ± 标准差，以及增减率的均值和标准差。

方法	每次试验的收集时间 [秒]			ID率[†a]
	1st	2nd	3rd	
ALL	159±47.4	176±51.7	213±71.9	0.400±0.516
不含 Hm	78.0±23.4	74.0±14.6	76.1±20.3	-0.0117±0.133
不含 CR	146±42.2	155±49.8	192±61.7	0.399±0.606
不含 ET	173±27.5	167±39.9	264±114	0.487±0.406

†a 表示将最后一次试验的收集时间除以第一次试验的收集时间所计算出的增减率。

不使用 Hm 的方法中，采集率根据相机视角的姿态而提高。

图 11 显示了用于拍摄通过 ALL 方法和不使用 Hm 的方法收集的 100 张图像的相机 6D 位姿散点图。该图还展示了一个二维边界框，它表示包含所有相机视点位置（绘制点）的最小长方体。图中绘制点的广泛分布表明数据集具有较大的多样性。

因此，似乎在未使用 Hm的情况下收集的数据集的多样性也很大。尽管如此，如果观察计算得到的距离 [m], , 与未使用Hm 的方法相比，ALL方法在所有试验中显示出更大的距离差异。表VII中所示的包含每个相机视点位置的二维边界框的两个计算体积[m^3] 相似，这是因为未使用Hm的方法中的均值距离更大。在 ALL方法中，显示视点的坐标系似乎相对密集地聚集在一起。这些结果可能是因为用户仔细地逐步移动相机，以捕捉所有可视化的矩形。相反，当半球被隐藏时，相机会随机移动。角距离计算为 $2 \times \arccos(\boldsymbol{q}^* \cdot \hat{\boldsymbol{q}})$。 $\boldsymbol{q}^*$ 和 $\hat{\boldsymbol{q}}$ 分别表示计算得到的四元数（指示相机姿态）和单位四元数（指示标记姿态）。计算得到的角距离为

表VII：计算了几种变异性，分别涉及从相机到物体的距离为 [m] 、包含所有位置的二维边界框的体积，以及所有视点中的角距离 [°] 。对于这两种度量，该表显示了均值 ± 标准差值。

方法	每次试验的距离 [m]		
	1st	2nd	3rd
ALL	0.762±0.0759	0.733±0.0729	0.738±0.0866
无 Hm	0.678±0.133	0.691±0.132	0.729±0.134
	体积 [m^3]		
方法	1st	2nd	3rd
ALL	0.595	0.534	0.747
不含 Hm	0.631	0.611	0.542
	每次试验的角距离 [°]		
方法	1st	2nd	3rd
ALL	152±16.8	156±19.7	157±14.7
不含 Hm	158±16.5	155±17.6	155±20.4

ALL 方法与不含 Hm 的方法之间具有可比性。

因此，半球可视化的效果使用户能够根据相机与物体位置之间的距离，采集多样化的多视角图像数据集。然而，无论 Hm 是否可见，多样性都不会发生显著变化，因为在所有情况下，都需要从不同视点捕捉场景，才能提高采集率。

VI. 注释结果评估

请注意，本评估并非本研究的主要范围。尽管我们将收集的数据集结果示例作为补充数据呈现，但这些结果不会影响 ALL 方法的有效性或无效性。

本节介绍了视觉系统的标注可行性和检测绩效。该系统使用一名用户收集的500张图像组成的数据集进行训练，该用户还另外收集了100张图像作为测试数据集。该用户



(a) 外观　(b) 几何形状

图12：用于标注可行性评估的目标物体。

与**第5节**所述实验的参与者不同。图12展示了本次评估使用的六个物体。我们从*NEDO*物品数据库[9]中选取了物体。为了评估绩效，我们计算了标注的交并比（IoU）、精确率、召回率和*F-score* ，以及目标检测的平均精确率（AP）和平均召回率（AR）。

A. 标注可行性

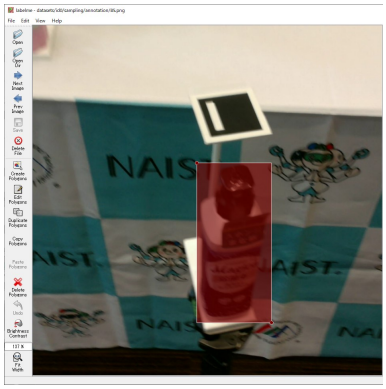
图 3**(b)** 展示了标注结果的可视化示例。为了使用四项指标评估 ALL 方法的标注结果，我们将其与人工标注的图像进行了比较。使用名为 *labelme*[10]，的人工标注工具，人们可以生成覆盖物体轮廓的边界框。人工标注员使用工具指定左上角和右下角，以创建边界框。人工标注 100 张图像平均需要 28 [分钟] ，而自动标注所需时间不到 1 [秒], ，表明时间绩效显著提升。

通过计算如图 13 所示的真正例 (TP)、假正例(FP) 和 假负例(FN) 值，我们获得了交并比、精确率、召回率和F分数。表 VIII 展示了标注中物体区域提取的结果。结果表明，所提出的自动标注与人工标注的结果之间存在一定程度的相似性，而非ALL 方法的准确率。该表展示了计算得到的交并比、精确率、召回率和 F 分数的均值 ± 标准差。

所有类别的四项指标均值分别为59.5%、60.5%、97.3% 和73.9%。召回率的均值被评为最高，且标准差小于其他指标的标准差。这些结果表明，人工标注中包含的相同区域很少被遗漏。此外，ALL方法生成的边界框大于人工标注生成的边界框。换句话说，所提出的标注中存在少量过度标注像素。分割像素的差异是由边界框生成过程造成的，该过程基于近似形状的大小。视觉标记的检测准确率为 100%；因此，根据

检测，但分割并非如此，均为 100%。因此，只要用户根据使用场景定义如何生成边界框，标注工作则依据用户的偏好进行。

[9]http://mprg.cs.chubu.ac.jp/NEDO DB/ [10]https://github.com/wkentaro/labelme



(a) 点击点



(b) 图像区域类别

图 13：人工标注。左图展示了使用工具（LabelMe）进行标注的示例。右图展示了分配给图像区域的类别，这些类别用于评估区域提取。

表 VIII：标注绩效。每个元素展示交并比 [%], 精确率 [%], 召回率 [%], 和 F 分数 [%] 的均值 ±标准差。这些数值表示两个值之间的相似度，而非准确率。

	指标 [%]			
ID[†a]	IoU	精确率	召回率	F 分数
1	68.9±7.19	70.4±7.92	97.6±3.97	81.4±5.09
2	62.2±10.3	63.7±10.1	95.7±10.8	76.1±9.60
3	58.7±9.83	59.1±10.2	99.2±1.52	73.5±7.67
4	49.3±11.9	50.5±13.6	96.5±11.1	65.2±11.4
5	61.3±10.9	62.5±11.1	96.0±10.1	75.3±10.2
6	56.4±9.66	56.9±10.5	98.9±2.18	71.7±7.39
Mean	59.5±5.96	60.5±6.15	97.3±1.36	73.9±4.89

[†a] ID 的数量与图12中所示的编号相对应。

B. 检测性能

本实验评估了使用通过 ALL方法收集的图像数据集训练的视觉系统。为了评估标注质量，我们在相同条件下，将开发的自动标注与人工标注进行了比较。我们使用基于深度神经网络模型的 EfficientDet D1 [36],作为物体检测器，并使用收集的数据集训练了该模型。我们确认了使用所提出的数据集和人工数据集训练的两个模型的检测结果。对于所提出的方法和人工方法，AP（50% < 交并比 < 95%）分别为74%和 78%。对于所提出的方法和人工方法，AR（50% < 交并比 < 95%）分别为78%和81%。这些结果表明，两个模型的准确率几乎相同。在本实验中，尽管我们仅为每个物体图像数据集收集了100张图像，但如果能够收集更多数据集，我们就可以使用具有更高复合系数（D2、D3、D4、D5 或 D6）的另一个 EfficientDet 模型，结果可能会得到改善。

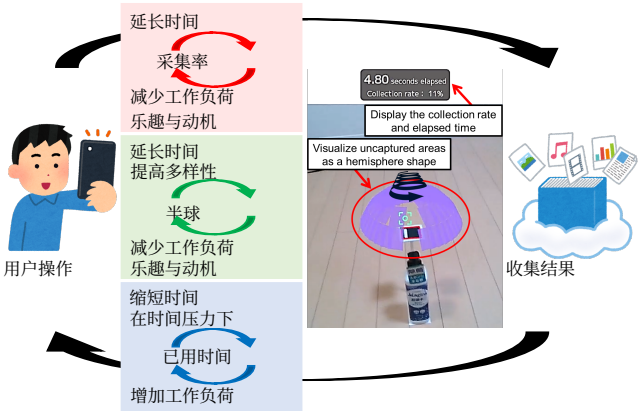


图 14：相互弥补彼此不足的应用功能之间的关系。

由于获得了与手动收集的数据集相同的绩效，因此所提出的 Web应用程序的有效性并未降低。

VII. 讨论

A. 数据集收集中的愉悦动机、效率与多样性

正如**第5.1节**所述结果所示，ALL 方法以及不使用 ET 的方法在时间需求方面得分较低（结果较好）。此外，不使用 Hm 的方法的用户绩效自我评价最低，这表明使用带有 Hm 的方法时，用户往往会对试验的质量给出更高评价。正如**第5.3节**所述，带有Hm 的方法可能会在其数据集采集试验中带来愉悦动机。在效率方面，正如**第5.4节**所述，不使用Hm的方法具有最短的收集时间；然而，其绩效自我评价较低，这表明如果花费适当的时间来收集数据集，用户可能会倾向于将自己的绩效估计得更高。此外，即使对于**第5.5节**中提到的数据集多样性，也最好显示半球。

综合考虑这些结果，显示收集状态（如Hm 和 CR）可能会导致耗时的收集试验。然而，结合可视化Hm、CR 和ET的 ALL 方法，会让用户产生愉悦感，并使他们能够更加专注于试验，而不是感受到过大的时间压力；因此，用户认为其试验质量更高。由于在采用 Hm的方法中多样性更好，其试验质量也更高，并且他们可能是在意识到进度的情况下进行收集试验的，将其视为关于收集状态的积极反馈。

在我们提出的系统中，收集试验期间 CR 和 Hm 的更新是表明目标取得进展的积极因素，而 ET 的增加则是表示绩效下降的负面指标。这可能是 ET 导致负荷较高的因素之一。为了平衡压力与绩效，有必要重新设计 ET，使其产生积极影响。例如，使用收集效率（件/秒）等指标，应当作为对用户的奖励。

总之，尽管与其他方法相比，收集时间无法大幅缩短，但所提出的应用程序能够以愉悦的方式激励用户，使用户能够在准确性和多样性方面收集高质量的数据集，并在适度的工作负荷（例如时间压力）下促进收集。图 14 展示了各应用程序功能之间相互弥补缺点的关系。由于即使是不具备 AR 技术使用经验或不了解训练数据集收集的参与者，其结果得分也往往与其他参与者相近，因此不了解相关知识的参与者也能从在线视觉反馈中受益，并高效地收集数据集。我们认为，实验结果的可靠性和普适性并不低。

B. 高度多样化的多视图数据集收集

在所收集数据集的多样性方面，仍有改进空间。在实验中，由于要收集的图像数量（可在图 1 所示的 *Maximum number ofimages* 中设置）被设定为 100，用户通过从不同方向拍摄目标物体，尽可能改变相机视角，而不是改变拍摄距离。未向用户提供具体指示。

未来，如果希望收集更多样化的数据集，虽然可以适当增加数量，但收集时间会更长，工作负荷也会相应增加。为解决这一问题，可以考虑实现其他模态功能，例如通过语音播报已用时间，以及通过音效报告收集进度。在未来的研究中，包括收集更多数据集的研究，我们将阐明这类多模态在线视觉反馈技术能否进一步提升愉悦感，从而增强参与动机。

C. 精确标注

在实验中，参与者使用了预先标定的相机。然而，对于 Web应用程序的一般用户而言，使用标记或棋盘格进行标定可能较为困难。因此，值得考虑采用相机自标定方法 [37]，这类方法无需标记或棋盘格即可估计相机的内参。

例如，COLMAP [38] 是一款运动恢复结构应用程序，提供相机自标定选项，只需使用多张图像即可轻松估计相机的内参。

VIII. 结论

本研究提出了一种基于Web应用程序的数据集采集系统，以避免繁琐且单调的标注过程。我们提出在应用程序的显示屏上显示三种数据集收集状态，包括将未采集区域显示为半球、收集

速率和已用时间，以更愉快的方式收集多视角图像数据集，从而减轻心理工作负荷和时间压力并提升动力。

用户测试结果揭示了所提出功能带来的有益和不利影响。总而言之，如图 14 所示，尽管显示半球和采集率往往会延长收集时间，但结合使用这些功能有望降低心理工作负荷，并让用户乐在其中。显示已用时间，可以通过在时间压力下进行数据集收集来缩短收集时间。然而，通过显示半球和采集率，为用户提供愉悦且能激发积极性的操作方式，可以减轻时间压力造成的心理工作负荷。此外，我们发现，半球显示增加了数据集的多样性；用户对自身绩效的估计较高，也因此获得了满足感。这些结果表明，结合所提出的在线视觉反馈功能，可能有助于提高动机，并平衡用户的工作负荷与绩效。这些发现与表 I 中提出的假设基本一致。

此外，使用所开发的数据集采集系统开展的实验表明，标注和目标检测具有可行性。未来的研究将推动开发一种简单而有效的多模态在线视觉反馈应用，以便为多种目标物体高效采集多样性丰富的多视角图像数据集。

参考文献

[1] M.Fujita, Y. Domae, A. Noda, G. A. G. Ricardez, T. Nagatani, A. Zeng,S.Song, A.Rodriguez, A. Causo,I. M. Chen,和T.Oga-sawara，“料箱拣选的重要技术是什么？基于一组绩效指标的竞赛机器人技术分析”，《先进机器人学》，第34卷，第7–8期，第560–574页，2020年。 [2] T. Kiyokawa, J.Takamatsu,和 S.Koyanaka，“混合工业废弃物未来机器人分拣机面临的挑战：一项综述”，《*IEEE*自动化科学与工程汇刊》，第21卷，第1023–1040页，2024年。[3] H. Zhang, L. Z. Liu, H. Xie,Y. Jiang, J. Zhou, 和 Y. Wang，“基于深度学习的机器人视觉：智能制造的高端工具”，《*IEEE*仪器与测量杂志》，第25卷，第2期，第27–35页，2022年。 [4] M. Naphade, S. Wang, D.C.Anastasiu, Z. Tang, M. Chang, Y. Yao, L. Zheng, M.S.Rahman,A.Venkatachalapathy,A.Sharma, Q. Feng, V. Ablavsky, S.Sclaroff, P. Chakraborty, A.Li, S. Li, 和 R. Chel- lappa，“第六届人工智能城市挑战赛”，*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）研讨会，第3346–3355页，2022年。 [5] T. Kiyokawa,K. Tomochika, J.Takamatsu,和 T. Ogasawara，“使用噪声屏蔽视觉标记实现基于深度学习的目标检测的全自动标注”，*IEEE RA-L*，第4卷，第2期，第1972– 1977页，2019年。 [6] M.Suchi, T. Patten, D. Fischinger, 和 M. Vincze，“EasyLabel：用于创建机器人RGB-D数据集的半自动逐像素目标标注工具”，*IEEE*国际机器人与自动化会议（*ICRA*）论文集，第6678–6684页，2019年。 [7] T. Kiyokawa,K. Tomochika, J.Takamatsu,和 T.Ogasawara，“利用后减弱多重视觉标记，高效收集并自动标注真实世界物体图像”，《先进机器人学》，第33卷，第24期，第1264–1280页，2019年。[8] D. De Gregorio, A. Tonioni, G.Palli, 和L. Di Stefano，“机器人技术中基于深度学习的半自动标注”，《*IEEE*自动化科学与工程汇刊》，第17卷，第2期，第611–620页，2020年。 [9] Y.Ishida 和H.Tamukoh，“用于目标检测与识别的半自动数据集生成及其在家用服务机器人上的评估”，机器人与机电学报，第32卷，第1期，第245–253页，2020年。

[10] I. Uygur、R. Miyagusuku、S. Pathak、H. Asama 和 A. Yamashita，《基于目标检测的数据融合与稀疏语义定位》，机器人与机电学报，第36卷第2期，第375–387页，2024年。[11] C. Agnew、C. Eising、P. Denny、A. Scanlan、P. Van De Ven 和 E. M. Grua，《量化真实标注质量对目标检测和实例分割绩效的影响》，IEEE Access，第11卷，第25 174–25 188页，2023年。[12] J. Murrugarra-Llerena、L. N. Kirsten 和 C. R. Jung，《我们能否信任目标检测的边界框标注？》，IEEE/CVF计算机视觉与模式识别会议（CVPR）研讨会论文集，第4813–4822页，2022年。[13] J. Ma、Y. Ushiku 和 M. Sagara，《提升标注质量对目标检测数据集的影响：一项初步研究》，IEEE/CVF计算机视觉与模式识别会议（CVPR）研讨会论文集，第4850–4859页，2022年。[14] H. Su、J. Deng 和 L. Fei-Fei，《面向视觉目标检测的众包标注》，AAAI人类计算研讨会，2012年。[15] B. Russell、A. Torralba、K. Murphy 和 W. Freeman，《LabelMe：用于图像标注的数据库和网络工具》，《国际计算机视觉杂志》（IJCV），第77卷，第157–173页，2008年。[16] R. Benenson、S. Popov 和 V. Ferrari，《由人工标注员参与的大规模交互式目标分割》，IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第11 700–11 709页，2019年。[17] C. Vondrick、D. Patterson 和 D. Ramanan，《高效扩展众包视频标注》，《国际计算机视觉杂志》（IJCV），第101卷第1期，第184–204页，2013年。[18] A. Dutta 和 A. Zisserman，《用于图像、音频和视频的VIA标注软件》，ACM多媒体国际会议论文集，第2276–2279页，2019年。[19] T. Kiyokawa、H. Katayama、Y. Tatsuta、J. Takamatsu 和 T. Ogasawara，《采用灵巧操作和快速可训练检测器的机器人废弃物分拣器》，IEEE Access，第9卷，第124 616–124 631页，2021年。[20] S. G. Hart 和 L. E. Staveland，《NASA任务负荷指数（NASA-TLX）的开发：实证研究与理论研究的结果》，《心理学进展》，第52卷，第139–183页，1988年。[21] J. Brooke，《SUS：一种“快速而粗略”的可用性评估方法》，《工业领域的可用性评估》，第189页，1996年。[22] L. von Ahn 和 L. Dabbish，《通过电脑游戏标注图像》，SIGCHI人机交互系统会议论文集，第319–326页，2004年。[23] J. Deng、J. Krause 和 L. Fei-Fei，《面向细粒度识别的细粒度众包》，IEEE计算机视觉与模式识别会议（CVPR）论文集，第580–587页，2013年。[24] I. Kavasidis、C. Spampinato 和 D. Giordano，《通过玩在线游戏生成目标检测的真实标注：生产性游戏还是娱乐性工作？》，IEEE计算机视觉与模式识别会议（CVPR）研讨会，2013年。[25] A. Carlier、O. Marques 和 V. Charvillat，《Ask’nSeek：一款用于目标检测和标注的新游戏》，欧洲计算机视觉会议（ECCV）研讨会论文集，第249–258页，2012年。[26] G. Hu、W. P. Tay 和 Y. Wen，《云机器人：架构、挑战与应用》，IEEE网络，第26卷第3期，第21–28页，2012年。[27] B. Kehoe、S. Patil、P. Abbeel 和 K. Goldberg，《云机器人与自动化研究综述》，《IEEE自动化科学与工程汇刊》，第12卷第2期，第398–409页，2015年。[28] W. Chen、Y. Yaguchi、K. Naruse、Y. Watanobe、K. Nakamura 和 J. Ogawa，《云机器人中的机器人协作研究：架构与挑战》，IEEE Access，第6卷，第36 662–36 682页，2018年。[29] P. Li、B. DeRose、J. Mahler、J. A. Ojea、A. K. Tanwani 和 K. Goldberg，《Dex-Net即服务（DNaaS）：基于云的稳健机器人抓取规划系统》，IEEE国际自动化科学与工程会议（CASE）论文集，第1420–1427页，2018年。[30] U. Klank、M. Z. Zia 和 M. Beetz，《从互联网数据库中选择三维模型以用于机器人视觉》，IEEE国际机器人与自动化会议（ICRA）论文集，第2406–2411页，2009年。[31] H. Bistry 和 J. Zhang，《利用智能相机系统处理复杂机器人视觉任务的云计算方法》，IEEE/RSJ智能机器人与系统国际会议（IROS）论文集，第3195–3200页，2010年。[32] E. B. Saff 和 A. B. J. Kuijlaars，《在球面上分布大量点》，《数学情报员》，第19卷，第5–11页，1997年。[33] N. Shirakura、T. Kiyokawa、H. Kumamoto、J. Takamatsu 和 T. Ogasawara，《联合使用无人机-无人潜航器收集海洋垃圾，并配备便于操作的图形用户界面》，IEEE Access，第9卷，67 443，2021年 67 432页。

[34] O. J. Dunn，《使用秩和进行多重比较》，*Technometrics*，第6卷，第241–252页，1964年。[35] B. A. Myers，《完成百分比进度指示器对人机界面的重要性》，*SIGCHI*人机交互系统会议论文集，第11–17页，1985年。[36] M. Tan、R. Pang 和 Q. V. Le，《EfficientDet：可扩展且高效的目标检测》，*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，第10 778–10 787页，2020年。[37] E. E. Hemayed，《相机自标定综述》，载于*IEEE*高级视频与信号监控会议论文集，2003年，第351–357页。[38] J. L. Sch¨onberger 和 J.-M. Frahm，《重新审视运动恢复结构》，载于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2016年，第4104–4113页。