CookAR: Affordance Augmentations in Wearable AR to Support Kitchen Tool Interactions for People with Low Vision

Jaewook Lee
University of Washington
Seattle, WA, USA

Andrew D. Tjahjadi
University of Washington
Seattle, WA, USA

Jiho Kim
University of Washington
Seattle, WA, USA

Junpu Yu
University of Washington
Seattle, WA, USA

Minji Park
Sungkyunkwan University
Suwon, Korea

Yapeng Tian
University of Texas at Dallas
Richardson, TX, USA

Jiawen Zhang
University of Washington
Seattle, WA, USA

Yuhang Zhao
University of Wisconsin-Madison
Madison, WI, USA

Figure 1: CookAR provides real-time object affordance augmentations in head-mounted AR to support cooking interactions. (A) a low vision participant uses CookAR to locate and grab a spoon; (B) the view in CookAR where kitchen tool affordances (grabbable vs. hazardous areas) are recognized and augmented by solid-colored overlays, with green overlays for graspable areas such as handles and red for hazardous areas such as a knife blade or the hot part of a tea kettle.

ABSTRACT

Cooking is a central activity of daily living, supporting independence and both mental and physical health. However, prior work has highlighted key barriers for people with low vision (LV) to cook, particularly around safely interacting with cooking tools, such as sharp knives or hot pans. Drawing on recent advancements in computer vision (CV), we present CookAR, a head-mounted AR system with real-time object affordance augmentations to support safe and efficient interactions with kitchen tools. To design and implement CookAR, we manually collected and annotated the first egocentric dataset of kitchen tool affordances, fine-tuned an affordance segmentation model, and developed an AR system with a stereo camera to generate visual augmentations. To validate CookAR, we conducted a technical evaluation of our fine-tuned model as well as a qualitative lab study with 10 LV participants for suitable augmentation design. Our technical evaluation demonstrates that our model outperforms the baseline on our tool affordance dataset, while our user study indicates a preference for affordance augmentations over the traditional whole object augmentations. Code is available at: https://github.com/makeabilitylab/CookAR.

CCS CONCEPTS

• Human-centered computing → Mixed / augmented reality; Accessibility systems and tools; • Computing methodologies → Computer vision.

KEYWORDS

augmented reality, accessibility, affordance segmentation, visual augmentation

1 INTRODUCTION

Cooking is an essential activity of daily living, supporting independence [54, 56, 70] and both mental and physical health [3, 31, 53, 70]. However, cooking also involves significant visual tasks that can be challenging or dangerous for blind and low vision (BLV) people, especially when interacting with kitchen tools, such as sharp knives or hot pans [3, 31, 38, 39, 74].

Unlike those who are completely blind, people with low vision (LV)—vision loss that cannot be corrected using glasses or contact lenses [10]—often rely on their residual vision in daily activities and use different low vision tools to enhance visual information [68, 69]. With recent advancements in AI-powered augmented reality (AR) technology, researchers have explored new possibilities for supporting LV individuals by automatically recognizing their environment and providing appropriate visual augmentations. Prior AR research prototypes have been developed to assist with tasks such as visual search [82], stair navigation [78], and sports [37]. However, none have been specifically designed to support meal preparation. Moreover, prior AR systems mainly focus on understanding effective augmentation designs [78], often oversimplifying the computer vision (CV) recognition in their development, thus neglecting the effects of technological limitations (e.g., CV inaccuracies and system delays) on user experience.

We introduce CookAR, a fully-functional wearable stereo AR prototype that recognizes and augments the affordances of cooking tools in real-time to support LV individuals with efficient and safe interactions in the kitchen. In contrast to prior research that augments objects as a whole [16, 82], we distinguish and augment the object affordance (i.e., component parts that afford interactions), such as the safe-to-handle “grabbable” areas and the dangerous-to-touch “hazardous” areas (Figure 1). To enable accurate affordance recognition, we constructed a custom egocentric image dataset for kitchen tool affordances by selecting and labeling images from the Epic Kitchens dataset [11] and fine-tuned an affordance segmentation model. We then leveraged stereo depth estimation using the ZED Mini1 stereo camera and an Oculus Quest 22 headset with video passthrough to precisely overlay affordance augmentations on the 3D environment in real-time.

To evaluate CookAR, we conducted a technical evaluation of our fine-tuned model as well as a three-part qualitative lab study with 10 LV participants to evaluate their experiences with CookAR and solicit feedback about potential affordance augmentation designs. For the model assessment, we found that our fine-tuned affordance segmentation model (mAP of 46.3%) outperformed the base RTMDet [46] model (mAP of 12.3%) in accurate tool affordance recognition and segmentation. For the three-part user study, LV participants were first asked to locate and pickup cooking tools across three conditions: (1) without CookAR (baseline), (2) with CookAR displaying whole object augmentations, and (3) with CookAR displaying affordance augmentations. They then completed a free-form cooking task with CookAR (Part 2) and brainstormed desired augmentation designs using design probes (Part 3). Findings indicate that participants prefer affordance augmentations over whole object augmentations in a kitchen, as they enable faster understanding of an object’s spatial arrangement and safe interaction parts. Most participants preferred affordance augmentations consisting of green solid overlay on grabbable areas and red outlines on hazardous areas. Moreover, participants identified five additional tool affordances with desired augmentations, including entry (e.g., cup rim), exit (e.g., carafe spout), containment (e.g., cup base), intersection (e.g., knife blade on butter), and activation (e.g., carafe buttons) areas, all of which should be outlined in a contrasting color (e.g., black or white).

In summary, our research contributions include: (1) CookAR, a fully-functional AI-powered wearable AR prototype that augments kitchen tool affordances for low vision users to enable safe and efficient tool interactions; (2) an egocentric affordance dataset for kitchen tools and an accompanying fine-tuned affordance segmentation model; and (3) user study results with 10 LV participants that reveal user experiences with CookAR, preferences for augmentation designs, and five newly desired affordance areas.

2 RELATED WORK

Our work builds on prior formative research exploring low-vision cooking, wearable AR to enhance accessibility, and affordance segmentation.

2.1 Challenges in Low Vision Cooking

Low vision people face various challenges in activities of daily living, such as cooking [3, 38, 39, 73, 74], shopping [68], navigation [68, 80], and sports [37, 62]. Among these activities, cooking is an essential task for an independent and healthy life [70]. However, cooking also poses major barriers and safety concerns to blind and low vision (BLV) people as they need to interact with various ingredients and kitchen tools, such as sharp knives and hot pans [32, 33]. As a result, BLV people tend to eat more pre-processed food or frequently dine at restaurants, which can negatively impact their health [31, 55].

To better understand how BLV people engage in cooking tasks, prior work has conducted both interview and observational studies [38, 40, 73, 74]. For example, Jones et al. [31] surveyed 101 BLV participants in the U.K. about their shopping and cooking experiences, revealing that vision loss made cooking difficult and the difficulty level was correlated with the severity of visual impairments. Li et al. [38] analyzed 122 YouTube videos of BLV people preparing meals and interviewed 12 BLV participants about their cooking experiences. They identified several cooking-related challenges, such as utilizing cooking tools and tracking object dynamics in the kitchen. A follow-up contextual inquiry study [39] examined how BLV people recognize cookware and utensils and measure ingredients. The study identified essential cooking-related information to convey, such as position, safety, and orientation information.

Specifically for low-vision (LV) people, Wang et al. [73] conducted a contextual inquiry study, observing and comparing the

1https://store.stereolabs.com/products/zed-mini
cooking experiences between six LV participants and four blind participants. They found that while blind participants relied on touch, LV people extensively used their vision during cooking. However, compared to blind people, LV people felt less confident, less safe, and more tired and stressed during cooking. Moreover, they were less satisfied with the currently-available cooking tools than blind people, indicating a need for technology that considers low vision people’s unique needs. The study also identified key challenges faced by LV people, such as distinguishing objects with low contrast and safely interacting with dangerous kitchen tools. Wang et al. [74] further interviewed six low vision rehabilitation professionals to understand current training strategies and tools to support cooking. However, they highlighted that current solutions cannot fully address all cooking challenges faced by LV people. Our research fills this gap by building an AR system to support safe and efficient cooking tool interactions for low vision people by providing visual augmentations.

2.2 Using Wearable AR to Enhance Accessibility

In the fields of accessibility and HCI, wearable AR has been used to support people with diverse disabilities, for example, AR systems that caption and visualize speech and sounds for deaf or hard of hearing (DHH) individuals [15, 27–29, 52, 59, 61, 65], support hands-free interactions with screen displays for people with upper body motor impairments [47, 48, 51], offer speech support for people with aphasia [76], and provide social cue therapy for children with autism [75].

Within the low-vision aid context, head-worn AR devices can selectively enhance user's vision by interpreting their environment and tasks [1, 77]. For instance, prior research has developed AR systems that can capture real-time video feeds of the surroundings and apply image processing techniques [12, 49, 82] to enhance visual information, such as edge enhancement [26, 35, 50], scene recoloring based on distance [14, 24, 72], and pixel remapping for visual field loss [21, 42, 44, 60]. However, while these solutions are beneficial for simple tasks like reading [12, 25, 66], they still lack semantic understanding of the scene and cannot effectively support more complex activities involving object interactions, such as cooking [74].

More recently, researchers have combined AR and CV to develop scene-aware visual augmentations aimed at assisting LV people with more intricate activities like visual search [82], stair navigation [78], wayfinding [79], obstacle maneuver [16], button pressing [36], and sports [37]. Nonetheless, no prior systems have addressed the unique cooking challenge that involves dynamic tool interactions. Moreover, prior AR research for low vision mainly focuses on designing and evaluating suitable visual augmentations for low vision people. They tend to oversimplify the CV recognition in system development, for example, using QR codes [78, 82] or existing spatial mapping APIs [16, 79] to anchor augmentations to the real-world environment. As a result, they overlook the technical challenges in building a real-time AI-powered AR system, as well as the potential impact of the technological limitations (e.g., CV errors, system latency) on users’ experiences. Our research addresses this gap by developing CookAR, a fully-functional wearable AR system that recognizes and augments kitchen tool affordances in real-time to support safe and efficient interactions in a cooking scenario.

2.3 Affordance Segmentation

In contrast to most prior research that augments whole objects [16, 37, 82], our research focuses on recognizing and augmenting tool affordances. Affordance is traditionally defined as “the opportunities for actions that objects offer, relative to the user’s ability to perceive and act on them” [18–20]. Highlighting object affordances can effectively guide human attention and actions [17, 63, 71]. Despite the prominence of affordance segmentation in robotics [2, 7, 8, 57, 58] and computer vision [9, 43, 45], automatic affordance augmentation has received comparatively less attention and applications in the field of HCI. Notably, there exists a gap for an affordance dataset in the context of a kitchen environment, especially designed for the needs of individuals with low vision. To address this, we built an AR system that focuses on affordance segmentation and enhancement. We created a new dataset focused on the affordances of kitchen tools by selecting and annotating image frames from the egocentric Epic Kitchens dataset [11] and fine-tuned an instance segmentation model on our dataset.

3 SYSTEM IMPLEMENTATION

We designed and built CookAR, an innovative solution that integrates AR and CV technologies to improve object interactions for individuals with low vision. Unlike traditional enhancements that target objects as a whole, our prototype highlights their affordances (i.e., functional parts), facilitating easier identification and interaction with areas to grasp or avoid. To create a fully-functional wearable AR system, we needed to address both the computer vision problem of accurately recognizing object affordances in real-time and the HCI problem of designing and rendering suitable affordance augmentations. In this section, we describe our approach for each step, including (1) collecting and annotating a dataset focused on the affordances of kitchen tools; (2) fine-tuning an instance segmentation and recognition model on our dataset to detect these affordances; and (3) building a head-mounted AR system with a stereo camera to render visual augmentations based on the outputs of our affordance recognition model.

3.1 Data Collection and Annotation

To train an effective AI model for affordance segmentation, we first need an appropriate dataset. However, to the best of our knowledge, there is no prior cooking tool dataset with annotations to enable affordance segmentation and recognition. Below, we describe our multi-step process to collect and annotate object affordances in egocentric cooking images. This labeled dataset is one contribution of our work and will be open-sourced upon paper acceptance to enable future research.

Data Collection. To build a kitchen tool affordance image dataset, we used an egocentric video dataset called Epic Kitchens [11], which consists of 100 hours of video footage of sighted people cooking in their homes. We selected the Epic Kitchens video dataset since it not only involves a wide range of cooking scenarios with
various kitchen tools, but also captures video feeds from the first-person perspective, which aligns with the egocentric nature of head-worn AR devices.

Because the Epic dataset is so large, we needed to filter critical frames of interest. We used YOLOv8 [30] trained on the MS COCO dataset [40] to detect and collect frames featuring cooking-related objects, such as spoons, knives, forks, cups, scissors, sinks, and dining tables. We skipped 20 frames after finding at least one of those objects to minimize repetition. We then manually reviewed the selected frames to empirically remove similar, excessively blurry, or irrelevant images, resulting in 4,928 key frames.

Data Annotation & Augmentation. We then labeled these frames using the image annotation feature in Roboflow, a platform that facilitates building and deploying end-to-end computer vision models by offering tools for annotation, training, and optimization. Roboflow enables labeling automation using the Segment Anything model (SAM) [34], allowing us to easily select and segment the interactive parts of objects (e.g., knife blade vs. knife handle) and add corresponding class labels.

Drawing on prior work in low-vision cooking [38, 39, 74], the research team collectively decided on 18 distinct classes: Knife Blade, Knife Handle, Spoon Bowl, Spoon Handle, Fork Tines, Fork Handle, Scissor Blade, Scissor Handle, Ladle Bowl, Ladle Handle, Spatula Head, Spatula Handle, Pan Base, Pan Handle, Cup Base, Cup Handle, Carafe Base, and Carafe Handle (Figure 2). When labeling, we adhered to the following heuristic: (1) the object should visually resemble the class it is being labeled as; and (2) the object should serve functions similar to those of the label class. For instance, a large wooden spoon can be tricky to label, as it can resemble a spoon, ladle, or spatula, and have versatile use such as stirring a pan (like a spoon), scooping contents from a pot (like a ladle), or lifting eggs (like a spatula) across different images in the dataset. We labeled these ambiguous objects based on their shape and use in a given frame. The annotations were cross-checked between six members of the research team to reduce errors and bias.

After annotating, we used various image augmentation techniques available on Roboflow to enhance the dataset for better generalizability across real-world scenarios, including: cropping with 0% minimum zoom and 40% maximum zoom, rotation between $-15^\circ$ and $+15^\circ$, brightness between -15% and +15%, blur up to 2.5px, and noise up to 0.1% of pixels. We then adjusted the images to fit a 640x480 resolution (i.e., the MS COCO average image resolution [40]) to accommodate our chosen base model’s preferences and facilitate their use in future research. This resulted in a final dataset of 10,152 images. The entire dataset is publicly available here: https://github.com/makeabilitylab/CookAR.

Figure 2: Roboflow annotation examples for each object class in our dataset.
With our fine-tuned RTMDet-Ins-l-Cook model, we built CookAR, a wearable AR prototype that can recognize and visually augment the affordances of kitchen tools in real-time. To implement CookAR, we needed to address key technical and HCI challenges, including: (1) how to spatially highlight object affordances in 3D space; (2) how to best visually indicate affordances to LV users to support but not overwhelm their existing perceptual vision systems.

To generate visual augmentations that align with the object parts in a 3D space, we built a custom stereo video see-through AR system by combining the ZED Mini stereo camera with an Oculus Quest 2 VR headset, as off-the-shelf AR headsets (e.g., Microsoft HoloLens 2) do not yet support long range real-time depth sensing and heavy CV models. The affordance visualizations themselves are rendered as colored polygon overlays. Specifically, we used green (hexcode #3BE8B0) to indicate a graspable area and red (#FC626A) to indicate a risky area. See Figure 1. As our user study section describes (Section 5), we also further explored and brainstormed other affordance augmentation designs.

Because our real-time CV model is computationally expensive, CookAR is tethered to a laptop with a NVIDIA 4080 mobile GPU. The CookAR system first captures image frames using the ZED Mini stereo camera and streams them to an external server via the Transmission Control Protocol (TCP) for affordance segmentation by our RTMDet-Ins-l-Cook model (confidence threshold of 0.4). Then, it converts the resulting JSON with affordance masks and labels into a Protocol Buffers message \(^5\) for efficient streaming. This message is then sent back to be processed by the ZED Mini API \(^6\), which deserializes the message back into a JSON and creates a ZED-compatible texture (or colored overlay) for each affordance mask.

Finally, the ZED Mini performs stereo depth estimation and overlays the textures onto the left and right image frames for binocular vision in the Quest 2 headset. While the ZED Mini stereo camera allowed us to prototype CookAR, it required our system to be tethered to a computer, which we alleviated by using lengthy cables.

In our latency analysis, we ran CookAR for five minutes and computed the average latency of each component: video streaming from ZED to computer took 16.76ms; affordance recognition took 15.95ms; result streaming back to ZED took 10.43ms; and depth estimation and augmentation rendering took 3.39ms. All other components had negligible impact on runtime. The overall latency is on average 46.82ms per frame (≈21.36 FPS), resulting in a near real-time system. System structure is summarized in Figure 3.

### 3.2 Model Fine-Tuning

To provide real-time object affordance information to LV users, we fine-tuned the RTMDet model \(^4\), specifically its RTMDet-Ins-l variant, on our kitchen tool affordance dataset. This model is the current state-of-the-art in real-time instance segmentation, \(^3\) offering robust accuracy and 300+ FPS on an NVIDIA 3090 GPU. RTMDet features large kernel depthwise convolution and batch normalization layers, pre-trained on MS COCO.

We opted to fine-tune RTMDet instead of training it from scratch, which allowed us to better leverage our smaller, class-specific dataset. To achieve this, we initialized the base RTMDet-Ins-l model with pre-trained weights, froze its backbone, and adjusted the model configuration file for our label classes, before training it on our kitchen tool affordance dataset. This customized model, dubbed RTMDet-Ins-l-Cook, was trained over 150 epochs with a batch size of 4 on a single CUDA-enabled NVIDIA 4080 GPU. Because RTMDet-Ins-l-Cook underwent fine-tuning on a dataset with affordance annotations, it can mimic an affordance segmentation model’s capabilities while retaining RTMDet’s real-time performance. We provide a technical evaluation in Section 4.

### 3.3 The CookAR Prototype

With our fine-tuned RTMDet-Ins-l-Cook model, we built CookAR, a wearable AR prototype that can recognize and visually augment the affordances of kitchen tools in real-time. To implement CookAR, we needed to address key technical and HCI challenges, including: (1) how to spatially highlight object affordances in 3D space; (2) how to run a real-time model capable of providing affordance feedback with minimal latency; (3) and how to best visually indicate affordances to LV users to support but not overwhelm their existing perceptual vision systems.

To generate visual augmentations that align with the object parts in a 3D space, we built a custom stereo video see-through AR system by combining the ZED Mini stereo camera with an Oculus Quest 2 VR headset, as off-the-shelf AR headsets (e.g., Microsoft HoloLens 2) do not yet support long range real-time depth sensing and heavy CV models. The affordance visualizations themselves are rendered as colored polygon overlays. Specifically, we used green (hexcode #3BE8B0) to indicate a graspable area and red (#FC626A) to indicate a risky area. See Figure 1. As our user study section describes (Section 5), we also further explored and brainstormed other affordance augmentation designs.

Because our real-time CV model is computationally expensive, CookAR is tethered to a laptop with a NVIDIA 4080 mobile GPU. The CookAR system first captures image frames using the ZED Mini stereo camera and streams them to an external server via the Transmission Control Protocol (TCP) for affordance segmentation by our RTMDet-Ins-l-Cook model (confidence threshold of 0.4). Then, it converts the resulting JSON with affordance masks and labels into a Protocol Buffers message \(^5\) for efficient streaming. This message is then sent back to be processed by the ZED Mini API \(^6\), which deserializes the message back into a JSON and creates a ZED-compatible texture (or colored overlay) for each affordance mask.

Finally, the ZED Mini performs stereo depth estimation and overlays the textures onto the left and right image frames for binocular vision in the Quest 2 headset. While the ZED Mini stereo camera allowed us to prototype CookAR, it required our system to be tethered to a computer, which we alleviated by using lengthy cables.

In our latency analysis, we ran CookAR for five minutes and computed the average latency of each component: video streaming from ZED to computer took 16.76ms; affordance recognition took 15.95ms; result streaming back to ZED took 10.43ms; and depth estimation and augmentation rendering took 3.39ms. All other components had negligible impact on runtime. The overall latency is on average 46.82ms per frame (≈21.36 FPS), resulting in a near real-time system. System structure is summarized in Figure 3.

### 4 TECHNICAL EVALUATION

We first conducted a technical evaluation of our fine-tuned RTMDet-Ins-l-Cook model, comparing its performance against the base RTMDet-Ins-l model on our kitchen tool affordance dataset. Findings indicate that our fine-tuned model is significantly more accurate in recognizing and segmenting affordances of cooking than the unmodified model.

### 4.1 Methods

To assess the performance of the base and fine-tuned models on our kitchen tool affordance dataset, we leveraged the model testing pipeline provided by the \(^\text{MMDetection library}^6\), a PyTorch-based open-source toolkit for object detection and segmentation, which

\(^{4}\) https://paperswithcode.com/sota/real-time-instance-segmentation-on-mscoco

\(^{5}\) https://protobuf.dev
performs evaluations using the test subset of a given dataset. Using Roboflow, we generated a testing set of 596 images with an 82-12-6 train-validation-test split and ensured that our fine-tuned model was not exposed to images in the test subset.

For instance segmentation tasks, accuracy is conventionally measured using three key metrics: segmentation mean average precision (mAP), AP at a 50% Intersection over Union (IoU) threshold (AP@50), and AP at a 75% IoU threshold (AP@75) [22]. Each metric serves a distinct purpose:

- **mAP** offers a comprehensive assessment of a model’s precision across various detection thresholds, by averaging precision at multiple recall levels for each class. It also aggregates results across a range of Intersection over Union (IoU) thresholds, from 0.5 to 0.95 in steps of 0.05, providing a holistic view of model performance across different degrees of overlap between the predicted masks and the ground truth;
- **AP@50** focuses on the precision of segmentation at a specific IoU threshold of 50%, a more lenient measure that considers a prediction correct if the overlap with the ground truth is at least half;
- **AP@75** evaluates precision at a stricter IoU threshold of 75%, demanding higher accuracy in the overlap between the predicted segmentation and the ground truth for a positive assessment.

IoU, central to these metrics, quantitatively evaluates the overlap between predicted segmentation masks and the actual ground truth, serving as a direct measure of accuracy in spatial alignment. We applied these metrics to compare the performance of our fine-tuned model against the baseline, aiming to capture the nuances of improvement across different levels of strictness in segmentation accuracy. We computed and compared the three metrics across the base and fine-tuned models.

### 4.2 Results

Our findings (Table 1) show that the base RTMDet-Ins-l model performs poorly on our affordance-specific dataset despite its competency on the COCO dataset [40] of 43.7% segmentation mAP. Conversely, our fine-tuned RTMDet-Ins-l-Cook model excelled in identifying and segmenting cooking tool components, demonstrating a significantly higher segmentation mAP of 46.3%, compared to the base model’s 12.3%. This improvement was also evident in our model’s performance at different IoU thresholds, with AP@50 and AP@75 reaching 74.9% and 48.6%, significantly outperforming the base model’s 19.9% and 13.2%. These results highlight the enhanced accuracy of our fine-tuned RTMDet-Ins-l-Cook model on our kitchen tool affordance dataset.

In addition, we show several inference results of RTMDet-Ins-l-Cook on images from the test subset of our dataset in Figure 4. Our model demonstrates impressive robustness, identifying and segmenting graspable, safe areas even when hands or other partial occlusions are present in the images.

### 5 USER STUDY

We then evaluated our CookAR prototype in a three-part qualitative lab study with 10 low vision participants. As initial work, our goals were to evaluate how LV participants might benefit from real-time object affordance augmentations to complete cooking tasks, solicit their reactions to a fully-functional but early-stage prototype (e.g., how do they react to augmentation errors), and co-brainstorm visual overlay designs via design probes. Participants provided feedback throughout the study and answered open-ended questions regarding their experiences, which were recorded and transcribed for later analysis.

#### 5.1 Participants

To achieve a diverse participant pool, we recruited 10 LV participants from two different cities via mailing lists and snowball sampling. Participants were screened using a demographic questionnaire, which gathered information on age, gender, vision condition, and prior experience with AR and AI technologies. The average age was 62.2 years (SD = 19.6), with a gender distribution of 70% female and 30% male. Participants had a broad range of low vision...
conditions with visual acuity ranging from 20/40 to 20/400 and visual field loss at different areas—see Table 2. Most participants reported little to no experience with AR and AI, except for P1 who had used both technologies.

5.2 Apparatus
The study was conducted in a well-lit lab environment. Participants sat in front of a large table, where we placed nine different kitchen tools—knife, spoon, fork, scissors, ladle, spatula, pan/pot, cup, and carafe. We used a dark green table cloth for the experiment table to produce low contrast. We also prepared a yellow wooden cutting board, a bowl, a piece of cheese, and a stick of butter for the participants to use in the study, however, CookAR can only recognize and augment the nine aforementioned kitchen tools at the current stage. Lastly, we recorded the experiment using a separate laptop and a smartphone on a tripod.

5.3 Procedure
The single-session 90-minute study consisted of three phases. In Part 1, we asked participants to grab cooking utensils in three conditions: (1) without CookAR, (2) with CookAR displaying augmentations for whole objects, and (3) with CookAR displaying affordance augmentations (Figure 5). In Part 2, participants completed a full cooking task where they made macaroni and cheese while using CookAR. Finally, in Part 3, participants brainstormed and identified potential applications of CookAR outside of kitchen contexts.

Part 2: Full Cooking Task. In Part 2 of the study, we asked participants to cook a macaroni and cheese dish using ARCook with affordance augmentations. We provided participants with step-by-step instructions to ensure that participants interact with all nine objects ARCook can recognize: (1) grab a cup of water and pour it into a carafe, (2) pour water into a pot using a carafe, (3) cut a piece of butter using a knife, (4) cut a piece of cheese using a pair of scissors, (5) put the macaroni, butter, and cheese into a pot, (6) stir with a spoon, (7) stir with a spatula, (8) place the finished macaroni and cheese in a bowl using a ladle, and (9) pick up a fork and enjoy. For the safety of our participants, we supplied a knife with a dull edge and avoided the use of heat. As participants completed this task, they were encouraged to think aloud, articulating how the affordance augmentations support or hinder their activities, how CookAR impacted their overall cooking experience, and any suggestions they had for augmentation designs. After completing the free-form cooking task, we asked participants to reflect on these same topics through open-ended questions.

Part 3: Brainstorming and Co-Design. In Part 3 of the study, we asked participants to brainstorm future designs and applications of CookAR. We first presented design probes (See Figure 6) and asked participants to critique them. The design probes illustrate various designs, including outlines which reduce visual clutter in comparison to solid-colored overlays, solely displaying either the grabbable or the hazardous augmentation, highlighting the more specific hazardous part such as the sharp edge of a knife blade rather than the whole blade, employing arrows to widen the area covered by the augmentations, and introducing a visual warning system when the user’s hand gets too close to a risky area. The design probes are inspired by prior work on low vision augmentations [16, 36, 78, 82]. Subsequently, we invited participants to propose augmentation design ideas for both simple objects like knives and more intricate objects with more interactive parts beyond grabbable and hazardous areas, like a carafe with many holes and buttons. Lastly, we asked participants to identify scenarios other than a kitchen where CookAR might be beneficial.

5.4 Analysis
We recorded participants’ quotes using Zoom. Transcriptions were first done by the video conferencing software, then the research team manually revised the transcripts. We collected 346 distinct quotes across our 10 LV participants, which we analyzed using reflexive thematic analysis [4, 5]. The first author, who facilitated every user study session, created an initial codebook by reviewing the revised study transcripts. The research team then collaboratively iterated on the codebook while checking for bias and coverage. With a final codebook consisting of 23 codes, the first author coded participants’ quotes, after which the team discussed the resulting themes. For likert-scale score analysis, we used a Wilcoxon signed-rank test since the data does not follow a normal distribution.

6 RESULTS
In our three-part qualitative study, participants completed tool-grasping tasks, a free-form cooking task, and a brainstorming session with design probes. Overall, participants found the real-time affordance augmentations helpful when interacting with various
<table>
<thead>
<tr>
<th>P#</th>
<th>Gender</th>
<th>Age</th>
<th>Left Eye Acuity</th>
<th>Right Eye Acuity</th>
<th>Description of Visual Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Male</td>
<td>30</td>
<td>No Light Perception</td>
<td>20/400</td>
<td>Coloboma dominates the right superior portion of my right eye.</td>
</tr>
<tr>
<td>P2</td>
<td>Female</td>
<td>83</td>
<td>20/200</td>
<td>20/100</td>
<td>Deteriorating eyesight from dry macular degeneration. Lost central vision on left eye. Central vision on the right eye is still there but not good. Have peripheral vision on both.</td>
</tr>
<tr>
<td>P3</td>
<td>Female</td>
<td>62</td>
<td>20/125</td>
<td>20/100</td>
<td>Low vision. Some holes in it, like black spots. Scar tissue.</td>
</tr>
<tr>
<td>P4</td>
<td>Male</td>
<td>65</td>
<td>20/20</td>
<td>20/60</td>
<td>Can see from 2/3’s of left eye, some far right peripheral vision from right eye.</td>
</tr>
<tr>
<td>P5</td>
<td>Female</td>
<td>70</td>
<td>20/200</td>
<td>20/100</td>
<td>Macular degeneration and side effects of chemotherapy. Blurry vision and need font enlargement to read. Visual field intact.</td>
</tr>
<tr>
<td>P7</td>
<td>Female</td>
<td>81</td>
<td>20/200</td>
<td>20/60</td>
<td>Diminished vision due to macular degeneration. Visual field intact.</td>
</tr>
<tr>
<td>P8</td>
<td>Female</td>
<td>80</td>
<td>20/50</td>
<td>20/50</td>
<td>Have dry macular degeneration with loss of some vision in the center of my left eye.</td>
</tr>
<tr>
<td>P9</td>
<td>Female</td>
<td>30</td>
<td>Light Perception</td>
<td>20/80</td>
<td>Can make out faces with right eye. Left eye is blind. Visual field intact.</td>
</tr>
<tr>
<td>P10</td>
<td>Female</td>
<td>71</td>
<td>20/60</td>
<td>20/100</td>
<td>I have Glaucoma. My field of vision is 5% eyesight. 5% in my left and 5% in the right remaining.</td>
</tr>
</tbody>
</table>

Table 2: Individual study participant information, including their gender, age, left and right eye acuity, and a self-reported description of their vision.

Figure 5: CookAR prototype with whole object augmentations (left) and affordance augmentations (right). The whole object augmentations are green instance segmentation masks, while the affordance augmentations are green (grabbable) and red (hazard) affordance segmentation masks.

cooking tools. They also suggested desired augmentation designs and key affordance parts. We expand on these findings below.

6.1 Affordance vs. Whole Object Augmentations

All but one participant (P6) preferred affordance augmentations over whole object augmentations for supporting kitchen tool interactions. They noted a trade-off between the augmentations’ utility and distraction, with the former generally outweighing the latter: “Seeing one color was less distracting than seeing two colors. But you’d have to know which end of the tool is the handle and which is the working end” (P2). We report participants’ feedback on the effectiveness and distraction of the augmentations below.

Effectiveness. In examining Likert data on perceived effectiveness, the findings showed no significant difference ($W = 36.5; p = 0.32$). However, participants on average gave higher ratings to affordance augmentations ($mean_w = 5.3; SD_w = 1.6$) over whole object augmentations ($mean_w = 4.6; SD_w = 1.4$). Affordance augmentations are advantageous in quickly understanding the overall scene (P1, P10), along with the placement and orientation of individual objects: “It helps to have two colors. I could see that more readily and quickly to understand how to use the object and how it is placed... your system is helpful because where the tool starts and ends and where the handle starts and ends is more clear” (P5). In addition, affordance augmentations become particularly useful when handling objects that have hazardous (9 out of 10 participants) (e.g., sharp, hot) or small (8/10) (e.g., door handles, buttons on appliances) parts, or have insufficient color contrast (7/10) (e.g., all silver or black cooking tools): “You’ve gotta show parts you can and shouldn’t grab. Green tells me that’s a safe place to go with my hand. Anything not green, I shouldn’t grab... It can help me avoid dangerous parts or perhaps even find small things like remote controllers” (P4). Furthermore, four participants expressed that for some objects with more complex interaction components than “just grab and don’t grab”
(P7), like a carafe with its handle, base, buttons, lid, and spout, they would accept the use of more than two colors, although “more than four colors can be quite distracting” (P7). We discuss augmentation designs in Section 6.3.

**Distraction & Comfort.** While participants qualitatively expressed that the whole object augmentations are less distracting than the affordance augmentations (6/10), the comfort and distraction Likert data was not significantly different ($W = 52.5; p = 0.88$ and $W = 46; p = 0.78$ respectively). The difference in average rating was also negligible, although participants on average found whole object augmentations to be slightly more comfortable ($\text{mean}_w = 5.1; \text{SD}_w = 1.3$ vs. $\text{mean}_a = 5.0; \text{SD}_a = 1.2$) and less distracting ($\text{mean}_w = 2.3; \text{SD}_w = 0.9$ vs. $\text{mean}_a = 2.5; \text{SD}_a = 1.1$). P6, who preferred whole object augmentations during the study, said “I think the more colors you have, the more distracting it becomes. So I prefer just the whole object in green than having 2 or 3 different colors. An outline would be better. I would definitely stay away from multicolored and just stick with one color. I can figure out its different parts.” Additionally, three participants shared that the whole object augmentations could be more useful depending on the scenarios. For tasks such as locating or avoiding objects, where interaction is not the goal, whole object augmentations are more preferable, since they are less distracting: “If I am looking for the remote controller, if it could make the remote stand more out in green or something, I don’t need its parts” (P3).

In summary, if a person’s intent is to interact with an object, affordance augmentations are more helpful than whole object augmentations. Conversely, in cases where interaction is not the objective, whole object augmentations may be preferred as they are less distracting. In our study, participants had to interact with various kitchen tools, hence most of them found our affordance augmentations more useful.

### 6.2 Experiences with CookAR

All participants were able to complete all free-form cooking steps within three to five minutes. However, due to current technical limitations in accurately segmenting object affordances, deploying heavy CV models, and rendering spatially accurate overlays in real-time on AR headsets, CookAR experienced recognition errors and perceived latency, which affected LV users’ experiences. All participants observed “flickering” and inaccurate affordance augmentations. Participants also pointed out that “the colors took some time to catch up” (P3) as they quickly rotated their head. Nonetheless, all but P6 saw potential in CookAR to assist with kitchen tool interactions and beyond: “I like the contrasting color. I just wish it more closely matched the object’s actual location. I think this highlighting scene is a great start. If the system is perfect, the dual color highlight system would be great and most useful. The system would be perfect if I am in a kitchen or just trying to grab really anything” (P1). P1, P5, P9, and P10 were particularly excited as they were able to better visually perceive object information: “This is fun! I can also use my eyes more to see shape and how it can be used. I want to try your system again in the future once you make it better” (P9).

Recognizing the potential of CookAR, participants identified additional use cases of a CookAR-like system that go beyond a cooking scenario, including cleaning (P3, P6, P7, P9), woodworking (P4, P5, P9), walking outdoors (P2, P6, P9), driving (P4, P6, P7), visiting a foreign kitchen (P2, P5), restaurants (P3, P9), gardening (P5, P10), watching sports (P7, P10), playing board games (P7, P10), going down stairs (P7, P9), identifying pill bottles (P7), and interacting with appliances with multiple buttons like a toaster (P1).

### 6.3 Desired Augmentation Designs

We report participants’ preferences on augmentation designs for grabbable and hazardous areas based on the design probes.

**Combining solid and outline augmentations.** As opposed to solid-colored overlays, nine participants preferred a mix of solid and outline augmentations, because solid colors are more salient, whereas outlines are less distracting: “Solid colors are helpful because they grab my attention... outlines are helpful because I can still see the part I’m trying to use with less distraction” (P5). Among those nine participants, all but P7 preferred solid-colored overlays for...
the grabbable area and outlines for the hazardous area because "the grabbable area is the most important" (P3, P4, P5, P8, P9), "all you need to know is its shape" (P4, P9), and "other parts should be outlined since you may want to do more with it, and solid color just makes it harder to use it" (P4, P8, P9). However, P7 preferred the outline for the grabbable area since it is not distracting and still shows the shape of the handle.

For the risky area to outline, P8 preferred highlighting solely the exact hazard (e.g., the sharp edge of the knife blade), as opposed to the entire dangerous part of an object (e.g., the whole knife blade), because she needs to know the relatively safer area for interaction. For example, as she described, "I might grab the top of the blade when I’m dicing or chopping. This tells you exactly where you shouldn’t touch." In contrast, all other participants wanted the outline augmentation because it is less distracting, yet still defines the overall shape of the hazardous area (P4, P5, P7, P9, P10) and what it is used for (P4, P7, P9). For example, P9 said, "I prefer to see the outlines on the [whole] blade, just so that way, you know which type of blade you’re grabbing. Cause a bread knife would look different from your knife. Some are thinner, some are fatter. People can be quite picky about their knife choices." P3, who preferred to outline dangerous parts, cautioned that "depending on how low vision you are on the spectrum, you may need solid colors to be able to see some of these objects."

Lastly, P1, P4, and P10 expressed concerns that overlaying perfectly aligned solid-colored affordance augmentations can be technically challenging. They suggested a colored circle may be enough, since they only need "a hint to see a glimmer of the object" and determine how the objects are oriented (P1).

Enhancing color contrast. Using colors to distinguish object affordances was well-received, as participants often color code their own cooking tools: "So I always try to get things color coded... especially if things are in drawers, it takes a lot of cognition for me to tell you what’s what. If it’s colored, it’s so much easier. This system is huge cause it’s doing color coding for me" (P1).

Every participant favored using green for safe-to-grab and red for dangerous areas, as "green signals ‘yes’ while ‘red’ signals no" (P4). However, P4, P7, and P9 struggled to clearly see our choice of red and requested a brighter shade of red, with P4 even suggesting white. Moreover, participants noted that the color contrast between the tool and the background is more important than the specific colors used, since a lot of objects in a kitchen are white, silver, or black with low contrast. For example, when cooking mac and cheese in the study, most (8/10) participants found it challenging to cut butter and understand where the yellow butter starts and ends because it was on a yellow wooden cutting board. To address this, P3, P4, P5, and P10 suggested the system should automatically select colors that contrast against the background: “The background you have it against will make a big difference, right? So on a darker background, I should be getting light colors” (P4). P7 jokingly said, “I mean, a green stick of butter could be weird, but it would at least let me cut better.”

Auditory feedback. Instead of visual augmentations, all participants preferred auditory feedback for warning in urgent scenarios (e.g., when the user’s hands get too close to a knife blade), as a visual warning could be easily missed by low vision users (P3, P5, P7, P9) and also makes the overall visual field busier (P4, P10). P3, P5, and P6 suggested small yet noticeable audio such as "beep beep," while P7 and P10 preferred explicit verbal waning (e.g., “stop”) since small noises can also be generated by other devices, such as a microwave or a fridge. P4 and P9 further suggested the system should employ different auditory signals for different hazards.

Action-aware Augmentations. As opposed to constantly augmenting all affordances, half of the participants suggested generating augmentations based on users’ current tasks or behaviors to reduce potential distraction. For example, with a knife, both the handle and blade can be augmented to start, then when a person grabs it, the handle augmentations could be turned off (P7); or, as a person gets close to a carafe with a cup of water, the rim of the carafe could be highlighted (P4). Moreover, seven out of 10 participants also suggested using voice commands to control the augmentations, such as turning on and off an augmentation or adjusting the augmentation design (e.g., colors or forms).

6.4 Additional Tool Affordances

In addition to the grabbable and hazardous affordances focused in our CookAR system, participants collectively suggested five other important affordances for kitchen tools: (1) entry area, (2) exit area, (3) containment area, (4) intersection area, and (5) activation area. We elaborate on these seven affordances along with participants’ preferred augmentation designs.

Grabbable area. A grabbable area is the part designed for safe handling or manipulation. This can include handles, grips, or any part intended for direct hand contact. Grabbable part of an object, participants preferred green solid-colored augmentations (Section 6.3).

Hazardous area. A hazard area is the part that poses potential risks or dangers to the user. This could include sharp edges, hot surfaces, or any part that can cause injury if touched or mishandled. For hazardous part of an object, participants preferred red outline augmentations (Section 6.3).

Entry area. An entry area is the part designed for initiating access, such as pouring. This could be the rim of a cup or a pot, the opening of a carafe, or any designated point that allows entry into an object’s containment space. All participants consistently noted that this area on an object should be augmented by an outline rather than a solid color, as the latter obstructs relevant actions like pouring or scooping: “The color blobs hide the item that you’re trying to put things into virtually completely. And so I can’t really tell if I am pouring something in correctly” (P4).

Exit area. An exit area is defined as the point through which contents are meant to be released. This could be the spouts, holes, or any defined pathway that guides content out of the object’s containment space, and it can be the same as the entry area for some objects, such as bowls and cups. Several participants (4/10) suggested that the carafe’s spout, similar to the entry area, should be outlined: “Highlighting the spout would be helpful if you had to pour, because if I poured in the wrong place, I wouldn’t know until something spills. I think I can pour more effectively if you highlighted this by aligning it with the edge of a cup or something. An outline would be great so I can see the water flowing out” (P7).

Containment area. A containment area has some depth and is meant to hold content within, such as food and liquid. This could
be the interior of a cup or pot, the base of a spoon or ladle, or any
defined space within the object that is meant to keep something
in. The current solid-colored overlays in CookAR interfere with
visibility of the containment space. Instead, all participants wanted
CookAR to augment only the entry and exit areas using outlines,
leaving the containment space without any augmentation.

Additionally, eight participants expressed that they need assis-
tance with understanding the depth of the containment space and
the amount of content it already holds. As P8 expressed, “A lot of
people with low vision cannot see inside and know how much water
they can pour. So somehow showing the water level and size of the
teapot is helpful. Mine is a lot bigger, it makes 12 cups or something,
and it’s all black, so it’s even harder to see what’s inside.” While P9
has a strategy to overcome this challenge by using her finger to
feel the liquid level, she cannot use it when the water is hot. She
thus suggested the system rendering “a blue disk” to indicate the
water level. In terms of augmenting the depth of the containment
area, participants suggested using a virtual line from the rim to
the bottom of the pot (P3, P4, P5, P9, P10), a measuring tape with
ticks (P4, P5, P9, P10), or a line with changing colors (e.g., a green
line that turns red as water fills up) (P4, P5). P10 further suggested
an auditory cue (e.g., a ‘ding’ sound) to indicate action milestones,
such as when reached quarter of a cup.

Intersection area. An intersection area is where parts of two
or more objects meet. This could be where a knife blade touches
the butter for cutting or where a cup touches a pot for pouring.
Interactions that require precise alignment between two objects
are particularly challenging to our LV participants. Half of the
participants suggested generating augmentations to highlight the
intersections or relationships between two interacting objects, for
example, the location where a knife cuts the butter (P5) or the
alignment between a ladle and a bowl when pouring (P9). As P9
mentioned, “Using a ladle has always been a problem for me. Pouring
the ladle into things is usually the hardest part, because you never
know if the ladle is in the right spot or too wide out of the way. Maybe,
if you have the ladle on top of a bowl, [CookAR should render] a
[virtual] shadow that gets casted onto the bowl.”

Activation area. An activation area is designed for initiating,
avivating, or turning on an object’s function or features. This could
be buttons, switches, touch-sensitive surfaces, or any interactive
components that trigger the operation of an object. Participants
identified activation areas on many household appliances, such as
buttons and dials on stove tops, microwaves, or coffee pots. They
are used for various purposes including starting a machine, opening
a lid, and adjusting settings. Seven participants preferred outline
augmentations for the activation area: “I just bought a vacuum with
multiple buttons. You would want different colors for the handles and
buttons” (P8). P10 further suggested a clock-like augmentation in
addition to the outline for turnable dials: “On a stove, I don’t know
what is medium heat. As I turn the knobs on a stove, the system
could show me ‘2 o’clock,’ ‘3 o’clock,’ and so on. ’6 o’clock’ is probably a
medium heat. ‘9 o’clock’ is probably a high heat.”

7 DISCUSSION

CookAR explores the use of affordance augmentations to enhance
kitchen tool interactions for LV people and advances the state-of
the-art in AI-powered AR systems. Results from our user study
indicates a preference for affordance augmentations over whole
object during tool interactions. Additionally, participants favored
augmentation designs that incorporate both solid-colored and out-
lined overlays with contrasting colors. We discuss design implica-
tions for affordance augmentations as well as current limitations
and future opportunities of AI-powered AR systems for low vision.

7.1 Design Implications for Affordances

Throughout the study, our LV participants proposed a range of
affordances for kitchen tools and indicated their preferred augmen-
tation designs for each. In this section, we summarize and expand
on these suggestions.

When to use affordance augmentations? Our study indicates
that visual augmentations should maximize utility and minimize
visual clutter and confusion. As such, it is critical to render aug-
mentations tailored to users’ intent and reduce distraction. For
example, affordance augmentations that involve multiple pieces
and colors would be more preferred to support interactions, while
basic whole-object augmentations are more suitable in general vis-
ual perception tasks such as avoiding obstacles. Beyond a kitchen,
affordance augmentations could also be applied to other scenarios,
as our participants suggested, such as gardening, playing board
games, and interacting with appliances (reaffirming [36]).

Where to apply affordance augmentations? Affordances can
refer to any object parts that indicate diverse action or interac-
tion opportunities. However, LV people face distinct interaction
challenges, resulting in unique affordances to augment for them.
In our qualitative study, we identified seven essential affordances
of kitchen tools that represent important yet challenging interac-
tion tasks for LV users. They include: (1) grabbable area, affording
touching and handling action; (2) hazardous area, affording risks
and avoidance; (3) entry, affording a target to aim at or pour in; (4)
exit, affording pouring out and usually requiring accurate align-
ment with the entry of another object (e.g., food transferring or
pouring); (5) containment area, affording holding content in and
preferring augmentations on the content amount (e.g., ingredient
measurement); (6) intersection area, affording touching or interac-
tion between two objects; and (7) activation area, affording control
features on an object. This affordance framework summarizes the
critical areas on objects as well as the hand-object (e.g., grabbable
vs. hazardous areas) and object-object (e.g., entry-exit alignment,
intersection between objects) relationships during interactions.

How to augment affordances? Different augmentations should
be designed for different types of affordances according to the in-
teraction tasks they indicate. Our study revealed the most preferred
augmentations for different affordances. Specifically, solid-colored
overlays with great visual emphasis are preferred for grabbable area
to enable fast perception and action, measuring augmentations (e.g.,
line with ticks) for containment area to indicate content amount,
and outlines for other affordances to avoid distraction and occlu-
sion. In terms of colors, augmentations should adopt colors with
high contrast against the environment. We also suggest leveraging
7.2 Challenges in AI-powered AR Development

This paper presents several key technical contributions across CV and HCI by constructing the first egocentric kitchen tool affordance dataset, fine-tuning an affordance segmentation model on our dataset, and developing a fully-functional stereo AR system that generates real-time affordance augmentations. Below, we reflect on key technical challenges stemming from both fields.

AI models for real-world use. Although our fine-tuned model outperforms the base model in accurately segmenting affordances of cooking tools, its mAP is still too low to successfully support dynamic activities like cooking in a real-world deployment study. For instance, its AP@75 is 48.6%, meaning in the worst case, about half of all predictions fail to achieve greater than 75% overlap with the ground truth affordance masks. The recognition results could become even worse in real-world use on a wearable AR system due to the users’ constant head motions and unique behaviors. For example, LV users tend to get much closer to objects than video data between the AR headset and an external server. However, the latency caused by the data transition prevented the overlays from keeping pace with the participants’ head motions, negatively impacting participants’ trust in CookAR’s intelligence and their perceptions of the system’s usability. To address the system latency, efforts are needed in both software development (e.g., real-time AI models) and hardware advancement (e.g., AR devices with powerful GPUs) to increase the usability of AR systems in dynamic real-world activities.

Affordance models and datasets. As opposed to object recognition models and datasets that attract significant attention in AI [13, 23, 40, 41], the research on affordance models and datasets remains nascent. To address this issue, we collected and labeled an affordance image dataset for kitchen tools and fine-tuned an object detection model on the affordance dataset to balance accuracy and efficiency for an in-lab user study but not a more naturalistic environment found in deployment studies. We suggest that when developing AI models, researchers should consider the potential real-world use cases, human needs, and integration to different hardware platforms (e.g., wearable AR) to enable use in practice.

7.3 Limitations & Future Directions

There are three primary limitations in this work. First, as an initial prototype, we conducted a qualitative user study with a relatively small number of participants to explore usability and solicit reactions to the idea of affordance augmentations in AR. Future work should conduct larger scale studies with more participants and diverse visual conditions. Second, the aforementioned technical challenges and limitations. Third and finally, the current CookAR system provides only one basic affordance augmentation—solid-colored overlays. Building upon the design insights in our study, future work should incorporate more augmentation options (e.g., outlines) and enable more flexible adjustment (e.g., colors, thickness of the outlines), to provide LV users more personalized experience.

8 CONCLUSION

In this paper, we introduce CookAR, a wearable AR prototype that can overlay affordance augmentations in real-time to support safe and efficient kitchen tool interactions for people with low vision. To build CookAR, we assembled an egocentric kitchen tool affordance dataset, fine-tuned an RTMDet-Ins-l model on our dataset (i.e., RTMDet-Ins-l-Cook), and leveraged this modified model with a stereo depth camera and an AR headset to generate real-time affordance augmentations in 3D space. In a technical evaluation of RTMDet-Ins-l-Cook, we found that it outperforms the base RTMDet-Ins-l model in accurately segmenting affordance masks of cooking tools. We then evaluated the CookAR prototype in a three-part lab study with LV participants. Findings indicate that participants prefer affordance augmentations over whole object augmentations for kitchen tool interactions. Participants expressed
a desire for an augmentation design with a combination of solid-colored and outlined overlays. The grabbable area should be augmented using solid green overlays, while other affordances including hazardous, entry, exit, containment, intersection, and activation areas should be outlined using a contrasting color. Participants also asked for enhanced customization and automatic adaptations for affordance augmentation designs. Our work suggests future designs of affordance augmentations to enhance kitchen tool interactions for LV people and advances the state-of-the-art in AI-powered AR experiences as low vision aids.

ACKNOWLEDGMENTS

This work was primarily supported by an NSF Graduate Research Fellowship and NSF CHS #1763199.

REFERENCES


