

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

Jiawen Zhang
University of Washington
Seattle, WA, USA

Jon E. Froehlich
University of Washington
Seattle, WA, USA

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

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

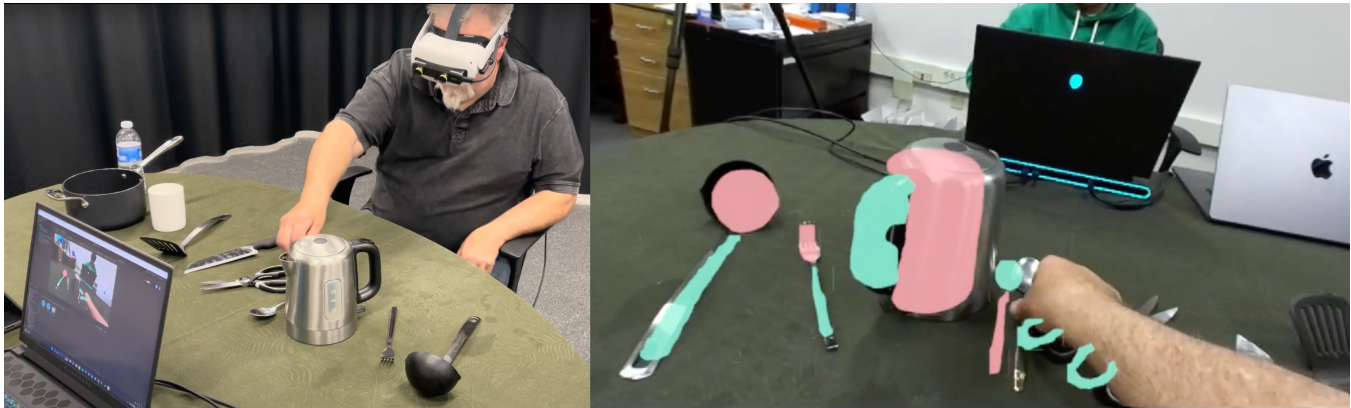


Figure 1: CookAR provides real-time object affordance augmentations in head-mounted AR to support cooking interactions. (A) a low vision participant uses CookAR to locate and grab a spoon; (B) the view in CookAR where kitchen tool affordances (grabbable vs. hazardous areas) are recognized and augmented by solid-colored overlays, with green overlays for grabbable 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 as well as mental and physical health. However, prior work has highlighted key barriers for people with low vision (LV) to cook, particularly around safely interacting with 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 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.

CCS CONCEPTS

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

KEYWORDS

augmented reality, accessibility, affordance segmentation, visual augmentation

ACM Reference Format:

Jaewook Lee, Andrew D. Tjahjadi, Jiho Kim, Junpu Yu, Minji Park, Jiawen Zhang, Jon E. Froehlich, Yapeng Tian, and Yuhang Zhao. 2024. CookAR:



This work is licensed under a Creative Commons Attribution International 4.0 License.

UIST '24, October 13–16, 2024, Pittsburgh, PA, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0628-8/24/10
<https://doi.org/10.1145/3654777.3676449>

Affordance Augmentations in Wearable AR to Support Kitchen Tool Interactions for People with Low Vision. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24), October 13–16, 2024, Pittsburgh, PA, USA*. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3654777.3676449>

1 INTRODUCTION

Cooking is an essential activity of daily living, supporting independence [54, 56, 70] as well as 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), researchers have explored new possibilities for supporting LV individuals by automatically recognizing their environment and providing appropriate visual augmentations, including stair navigation [78], visual search [82], and sports [37]. While promising, these previous AR systems focus primarily 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. Moreover, though there has been substantial formative work in BLV cooking within the HCI literature [38, 39, 74], no previous AR system has been built to address LV cooking specifically.

We introduce *CookAR*, a wearable stereo AR prototype that recognizes and augments cooking tool *affordances* in real-time to support LV meal preparation. In contrast to prior AR research that highlights objects as a whole [16, 82], we distinguish and augment the *object affordance* specifically (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 combined the *ZED Mini*¹ stereo camera and an *Oculus Quest 2*² headset to achieve a video passthrough AR system with CV and stereo depth estimation capabilities to precisely overlay affordance augmentations on the 3D environment in near 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. 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 tool affordance recognition and segmentation. For the three-part user study, LV participants were first asked to locate and pick up cooking tools across three conditions: (1) with their typical method in daily life such as wearing corrective lenses (i.e., real-world baseline); (2) with *CookAR* displaying whole object augmentations (i.e.,

augmentation baseline); 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 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. To enable others to build off our work, this dataset and model are open-sourced at: <https://github.com/makeabilitylab/CookAR>; 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 studies on low-vision cooking, wearable AR for accessibility, and affordance segmentation.

2.1 Challenges in Low Vision Cooking

People with low vision (LV) face challenges in everyday activities, such as cooking [3, 38, 39, 73, 74], shopping [68], navigation [68, 80], and sports [37, 62]. Among these, cooking is an essential task for an independent and healthy life [70]. However, this task also poses major accessibility barriers and safety concerns to BLV individuals, including interacting with sharp knives and hot pans [32, 33]. Consequently, they 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 that the level of difficulty was correlated to 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. This study highlighted essential cooking-related information to convey, such as position, safety, and orientation of objects.

Specifically for LV people, Wang *et al.* [73] conducted a contextual inquiry study, observing and comparing the cooking experiences between six LV participants and four blind participants. They found that while blind participants relied on touch, LV people

¹<https://store.stereolabs.com/products/zed-mini>

²<https://www.meta.com/quest/products/quest-2/>

used their vision extensively while cooking. However, compared to blind people, LV people felt less confident, less safe, and more tired and stressed due to their reliance on impaired vision. Moreover, LV individuals were less satisfied with existing cooking tools compared to blind people, highlighting the need for technology that considers their unique needs. The same study also identified key challenges LV people face, such as distinguishing objects with low contrast and safely interacting with dangerous kitchen tools. Wang *et al.* [74] further interviewed six LV rehabilitation professionals to understand current training strategies and tools for cooking. They emphasized that existing solutions are insufficient to overcome all cooking challenges LV people experience. Our research fills this gap by creating an AR system that supports LV people in safely and efficiently using cooking tools through visual augmentations.

2.2 Using Wearable AR to Enhance Accessibility

In accessibility and HCI, wearable AR has been used to support people with diverse disabilities. For example, AR glasses can caption and visualize speech and sounds for deaf or hard of hearing (DHH) people [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 users' vision by interpreting their environment and tasks [1, 77]. For instance, prior research has developed AR systems that can capture real-time video feeds 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 like cooking [74].

More recently, researchers have combined AR and CV to create scene-aware visual augmentations to assist LV people with more intricate activities like visual search [82], stair navigation [78], wayfinding [79], obstacle maneuvering [16], button pressing [36], and sports [37]. Nonetheless, no prior systems have addressed the unique challenges of cooking involving dynamic tool interactions. Moreover, prior AR research for low vision primarily focuses on designing and evaluating visual augmentations. Therefore, they tend to oversimplify CV recognition in system development, such as by using QR codes [78, 82] or existing spatial mapping APIs [16, 79] to anchor augmentations to the real world. This approach neglects the technical challenges of building a real-time AI-powered AR system and the potential impacts of technological limitations (*e.g.*, CV errors, system latency) on user experience.

Our research advances the field by introducing CookAR, a wearable AR system that recognizes and augments kitchen tool affordances in near real-time, enhancing safe and efficient interactions during cooking.

2.3 Affordance Segmentation

In contrast to most prior studies that augment objects as whole [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 is a lack of an egocentric affordance dataset specifically created for the needs of LV individuals. To address this gap, we first created a new dataset focused on kitchen tool affordances by selecting and annotating image frames from the egocentric *Epic Kitchens* video dataset [11] and fine-tuned an instance segmentation model. We then built an AR system that can segment and enhance tool affordance information.

3 SYSTEM IMPLEMENTATION

To support safe and efficient LV hand-object interactions with kitchen tools, we designed and built *CookAR*, a wearable stereo AR prototype that recognizes and augments cooking tool affordances in near real-time. Unlike traditional enhancements that target objects as a whole, our prototype highlights their affordances (*i.e.*, functional parts), facilitating identification and interactions 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 on the recognized tool affordances. The labeled dataset and fine-tuned model weight are one contribution of our work and are open-sourced at: <https://github.com/makeabilitylab/CookAR>.

3.1 Data Collection and Annotation

To train a CV model for affordance segmentation, we first needed a labeled dataset. However, to our knowledge, there is no prior cooking tool dataset with annotations to enable affordance segmentation. Below, we describe our multi-step process to collect and annotate object affordances in egocentric cooking images.

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

Because the *Epic Kitchens* dataset is large, we first needed to filter for 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. To minimize repetition, we skipped 20 frames after finding at least one of those objects. We then manually reviewed the selected frames to empirically remove similar, excessively blurry, or irrelevant images, resulting in 4,928 key frames.

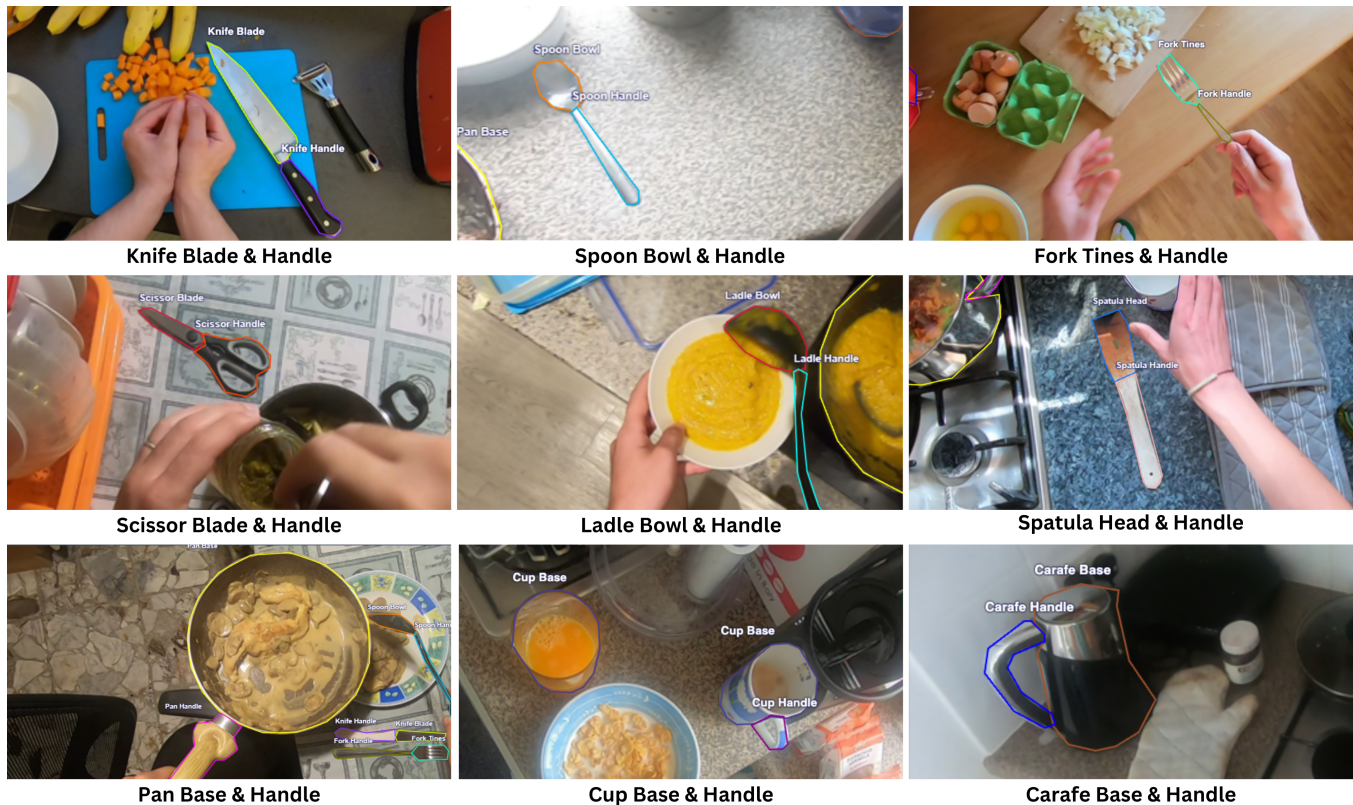


Figure 2: Example Roboflow annotations for each object class in our dataset (18 classes total).

Data Annotation & Augmentation. We then labeled these frames using *Roboflow*³, an online tool for annotating, training, and optimizing CV models. Roboflow also supports labeling automation using the *Segment Anything model* (SAM) [34], allowing us to easily select and segment interactive parts of objects (e.g., knife blade vs. knife handle) and add corresponding class labels.

Drawing on prior research [38, 39, 74], the research team identified 18 distinct classes of kitchen tools commonly used by BLV individuals: *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 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. Six research team members performed the annotations, each labeling a subset of images and having their work reviewed by another researcher 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° 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.

3.2 Model Fine-Tuning

To provide real-time object affordance information to LV users, we fine-tuned the *RTMDet* model [46], specifically its *RTMDet-Inst-l* variant, on our kitchen tool affordance dataset. This model is the current state-of-the-art in real-time instance segmentation⁴, 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 [40].

We opted to fine-tune *RTMDet* instead of training it from scratch, as this allowed us to better leverage our smaller, class-specific dataset. To achieve this, we leveraged the fine-tuning pipeline provided by the *MMDetection* library [6], a PyTorch-based open-source

³<https://roboflow.com>

⁴<https://paperswithcode.com/sota/real-time-instance-segmentation-on-mscoco>

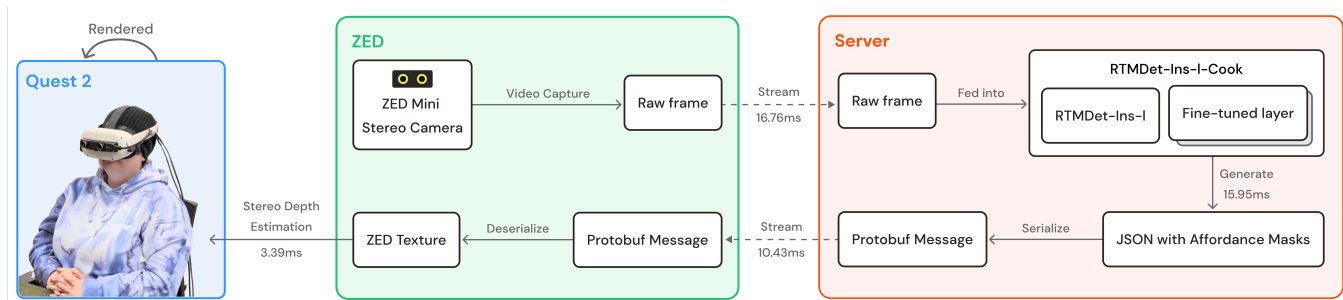


Figure 3: System overview of CookAR showing how data flows from the ZED Mini Stereo camera to an external camera for affordance segmentation, then sent back to ZED for rendering on the Quest 2 headset.

toolkit for object detection and segmentation, which supports various models including RTMDet. We initialized the base RTMDet-Ins-I 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-I-Cook*, was trained over 150 epochs with a batch size of 4 on a single CUDA-enabled NVIDIA 4080 GPU. Because RTMDet-Ins-I-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 of our model in Section 4. We also open-sourced our fine-tuning steps and code, giving researchers the tools to expand our dataset and fine-tune their own models: <https://github.com/makeabilitylab/mmdet-fine-tuning>.

3.3 The CookAR Prototype

With our RTMDet-Ins-I-Cook model, we built CookAR, a wearable AR prototype that can recognize and visually augment the affordances of kitchen tools in near real-time. To implement CookAR, we addressed three key technical and HCI challenges, including: (1) how to spatially highlight object affordances in 3D space; (2) how to develop a real-time system pipeline to provide affordance feedback with minimal latency; and (3) how to best visually indicate affordances to LV users to support but not overwhelm their existing visual perceptions. See video figure for a demonstration.

To generate visual augmentations that align with the object parts in 3D space (e.g., a knife grip, a cup handle), we built a custom stereo video see-through AR system by combining the *ZED Mini* stereo camera with an *Oculus Quest 2* VR headset. While off-the-shelf AR headsets such as the Microsoft HoloLens 2 may eliminate the need for a bulky video passthrough system, they do not yet support long range real-time depth sensing. We visualize the affordance representations as colored polygon overlays. Though the colors are settable, we currently use **green** (hexcode #3BE8B0) to indicate a graspable area and **red** (#FC626A) to indicate a risky area (Figure 1). As described in our user study section, we also further explored and brainstormed other affordances and 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-I-Cook model (confidence threshold of 0.4). Then, the server converts the resulting JSON with affordance masks and labels into a *Protocol Buffers* message⁵ for efficient streaming. This message is then sent back for processing by the ZED Mini API [67], 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. To enable participants to move freely during the study session, we connected the CookAR system to a computer using long (16 feet) cables.

In our latency analysis, we ran CookAR for five minutes and computed the average latency of each component: video streaming from ZED to the server 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. See Figure 3 for a system flow diagram.

4 TECHNICAL EVALUATION

We first conducted a technical evaluation of our RTMDet-Ins-I-Cook model, comparing its performance against the base RTMDet-Ins-I model on our kitchen tool affordance dataset. Findings indicate that our model is significantly more accurate in recognizing and segmenting affordances of cooking tools 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 used MMDetection’s [6] model testing pipeline, which performs evaluations using the test subset of a given dataset. With Roboflow, we generated a test set of 596 images with an 82-12-6 train-validation-test split and ensured that our 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]. *IoU*, central to these metrics, quantitatively evaluates the overlap between

⁵<https://protobuf.dev>



Figure 4: Example inferring results on images from the test subset of our dataset. These images demonstrate how the RTMDet-Ins-l-Cook identifies and segments graspable, safe areas—even in the presence of hands or other partial occlusions.

Model Name	mAP	AP@50	AP@75
RTMDet-Ins-l (on COCO)	0.437	0.660	0.470
RTMDet-Ins-l (on our dataset)	0.123	0.199	0.310
RTMDet-Ins-l-Cook (on our dataset)	0.463	0.749	0.486

Table 1: Affordance segmentation results. Our fine-tuned affordance segmentation model, *RTMDet-Ins-l-Cook*, achieves superior performance across all metrics on our kitchen tool affordance dataset, outperforming the state-of-the-art baseline *RTMDet-Ins-l* model. For reference, we also include *RTMDet-Ins-l* results on the COCO dataset.

predicted segmentation masks and the actual ground truth, serving as a direct measure of accuracy in spatial alignment. We explain each metric in detail below:

- *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* uses a precision of segmentation at the IoU threshold of 50%, a more lenient measure that considers a prediction correct if the overlap with 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 ground truth.

We applied these metrics to compare our model’s performance against the baseline, aiming to capture the nuances of improvement across different levels of strictness in segmentation accuracy.

4.2 Results

Our findings (Table 1) indicate that the base RTMDet-Ins-l model performs poorly on our affordance dataset despite its competency on the COCO dataset [40] with a 43.7% segmentation mAP. Moreover, 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%. In Figure 4, we show several example inference results of RTMDet-Ins-l-Cook on test images. Our model demonstrates impressive robustness, identifying and segmenting graspable, safe areas even when hands or other partial occlusions are present in the images. Overall, these findings highlight the enhanced accuracy of our RTMDet-Ins-l-Cook model.

5 USER STUDY

As a complement to our technical evaluation, we conducted a three-part qualitative lab study with 10 low vision (LV) participants. Our goals were threefold: first, to evaluate how LV participants might benefit from real-time object affordance augmentations when completing cooking tasks; second, to solicit their reactions to a fully-functional but early-stage prototype (*e.g.*, how do they react to augmentation errors); finally, to 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 cities (Madison, WI and Seattle, WA) 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 to simulate a visually challenging environment with 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, although CookAR can only recognize and augment the nine aforementioned kitchen

P#	Gender	Age	Left Eye Acuity	Right Eye Acuity	Description of Visual Field
P1	Male	30	No Light Perception	20/400	Coloboma dominates the right superior portion of my right eye.
P2	Female	83	20/200	20/100	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.
P3	Female	62	20/125	20/100	Low vision. Some holes in it, like black spots. Scar tissue.
P4	Male	65	20/20	20/60	Can see from 2/3's of left eye, some far right peripheral vision from right eye.
P5	Female	70	20/200	20/100	Macular degeneration and side effects of chemotherapy. Blurry vision and need font enlargement to read. Visual field intact.
P6	Male	50	20/40	No Light Perception	Blind in right eye. Need glasses for left. Visual field intact.
P7	Female	81	20/200	20/60	Diminished vision due to macular degeneration. Visual field intact.
P8	Female	80	20/50	20/50	Have dry macular degeneration with loss of some vision in the center of my left eye.
P9	Female	30	Light Perception	20/80	Can make out faces w/right eye. Left eye blind. Visual field intact.
P10	Female	71	20/60	20/100	I have Glaucoma. My field of vision is 5% eyesight. 5% in my left and 5% in the right remaining.

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

tools at the current stage. Lastly, we recorded the experiment using a 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 with CookAR and two baselines. In Part 2, participants completed a full cooking task where they made macaroni and cheese while using CookAR with affordance augmentations. Finally, in Part 3, participants brainstormed tool affordances and desired augmentation designs while examining and critiquing design probes. Prior to the study tasks, participants completed a tutorial to become familiar with CookAR. In total, participants interacted with our AR device for ~30 minutes across the study: a 5-minute tutorial, 10 minutes of object grabbing, and 15 minutes of free-form cooking. We provide more details below. The full study protocol is in the Appendix.

Tutorial. Participants first completed a tutorial task, where they interacted with a cooking pot using CookAR. Participants wore and adjusted the Oculus Quest 2 headset and freely explored CookAR and its affordance augmentations. Once they achieved a comfortable fit and an understanding of CookAR, the study proceeded.

Part 1: Tool Grabbing Task. In Part 1, participants were asked to locate and pick up cooking tools under three conditions: (1) real-world baseline (*i.e.*, without CookAR), (2) augmentation baseline (*i.e.*, whole object augmentations) [16, 82], and (3) CookAR (*i.e.*, affordance augmentations) (See Figure 5). We counterbalanced the condition order via *Latin Square*. In each condition, participants conducted five trials of picking-up tool tasks. We randomly chose a cooking tool per trial from the nine kitchen tools on the experiment table (See Section 5.2). Participants were asked to keep their eyes closed until the researcher named an object to reduce the effect of memory on task performance. Moreover, the researcher rearranged the placement and angle of the cooking tools between each condition. After each condition, we asked participants three 7-point Likert questions about effectiveness, comfort, and distraction, as

well as open-ended questions regarding their experience with each augmentation condition. After all 15 trials, we asked participants to compare the pros and cons of the three conditions, suggest improvements for CookAR, and identify potential applications of CookAR outside of kitchen contexts.

Part 2: Full Cooking Task. In Part 2, we asked participants to cook a macaroni and cheese dish using CookAR with affordance augmentations. We provided participants with step-by-step instructions for consistency and to ensure participants interacted with all nine objects CookAR 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 supported or hindered 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, we asked participants to brainstorm future designs and applications of CookAR. Drawing on prior work in low vision augmentation [16, 36, 78, 82], we created and presented design probes of various augmentation designs (Figure 6) and asked participants for feedback. The design probes included: (1) outlines to reduce visual clutter in comparison to solid-colored overlays, (2) solely displaying either the grabbable or the hazardous augmentation, (3) highlighting the more specific hazardous part such as the sharp edge of a knife blade rather than the whole blade, (4) employing arrows to widen the area covered by the augmentations, and (5) introducing a visual warning system when the user's hand gets too close to a risky area. After presenting

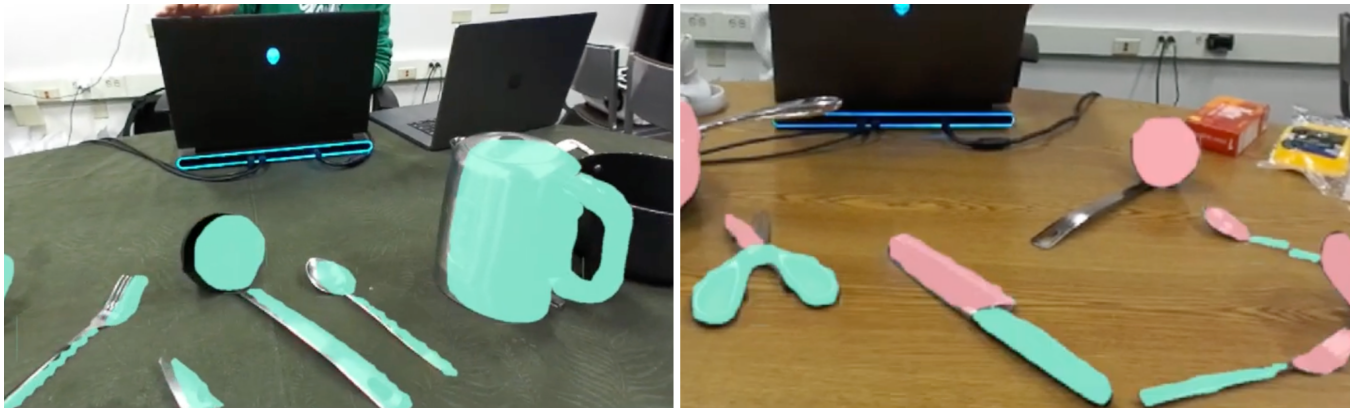


Figure 5: The CookAR prototype with whole object augmentations (left) and affordance augmentations (right). The whole object augmentations are green instance segmentation masks, while the affordance augmentations are green (grabbable) and red (hazard) affordance segmentation masks.

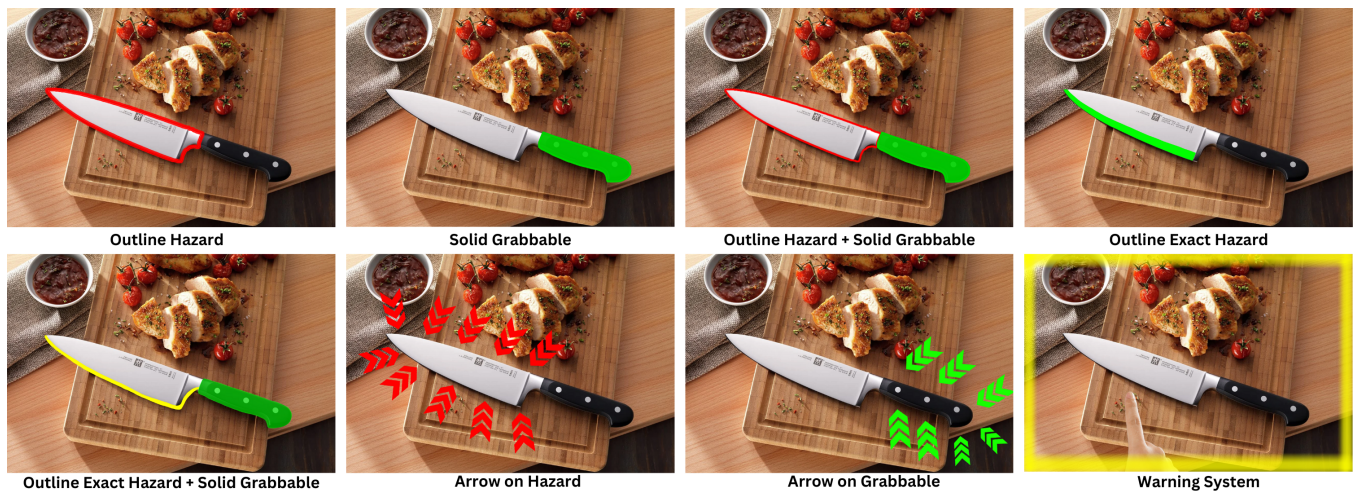


Figure 6: Design probes used in Part 3 of the study to spark design ideas.

our own designs, we invited participants to propose other ideas for both simple objects like knives as well as more complex objects with additional interactive parts beyond grabbable and hazardous areas, like a carafe, which has many openings and buttons with different purposes. Lastly, we asked participants to identify other scenarios where affordance augmentations 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 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 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 scores on perceived effectiveness, we found no significant difference between whole object augmentations and affordance augmentations ($W = 36.5$; $p = 0.32$). However, participants on average gave higher ratings to affordance augmentations ($mean_a = 5.3$; $SD_a = 1.6$) 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 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 additional augmentation designs in Section 6.3.

Distraction & Comfort. While most participants qualitatively expressed that the whole object augmentations are less distracting than the affordance augmentations (6/10), the comfort and distraction Likert scores were 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 ($mean_w = 5.1$; $SD_w = 1.3$ vs. $mean_a = 5.0$; $SD_a = 1.2$) and less distracting ($mean_w = 2.3$; $SD_w = 0.9$ vs. $mean_a = 2.5$; $SD_a = 1.1$). P6, who preferred whole object augmentations, 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, we found that if a person intends 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. Despite the positive qualitative feedback on CookAR, the Likert scores for the real-world baseline (without

CookAR) were higher than those for both the whole object augmentations (effectiveness: $W = 94$; $p < .001$, comfort: $W = 95$; $p < .001$, $W = 10$; distraction: $p < .001$) and affordance augmentations (effectiveness: $W = 77.5$; $p < .05$, comfort: $W = 52.5$; $p < .001$, distraction: $W = 10$; $p < .001$). This may be due to current limitations of AI-powered AR systems such as accuracy and latency. We discuss this further in Section 7.2.

6.2 Free-form Cooking with CookAR

All participants were able to complete all free-form cooking steps within three to five minutes. However, due to technical limitations in accurately segmenting affordances, deploying heavy CV models, and rendering spatially accurate overlays in real-time on AR headsets, participants experienced recognition errors and latency with CookAR. All participants observed “flickering” and inaccurate 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 [a tool] can be used. I want to try your system again once you make it better” (P9). Participants also identified the following additional use cases for a CookAR-like system: 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 less 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.”*

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 many kitchen objects 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 short yet noticeable audio such as *“beep beep,”* while P7 and P10 preferred explicit verbal warning (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 on by 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. For grabbable part of an object, participants preferred green solid-colored augmentations.

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.

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 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 assistance 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 water reaches 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, activating, 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. For example, P8 said: “I just bought a vacuum with multiple buttons. You would want different colors for the handles and buttons” Seven participants preferred outline augmentations for the activation area. Additionally, P10 further suggested a clock-like augmentation in addition to an 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 real-time 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 augmentations during tool interactions. Additionally, participants favored augmentation designs that incorporate both solid-colored and outlined overlays with contrasting colors. We discuss design implications 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, LV participants proposed a range of affordances for kitchen tools and indicated preferred augmentation designs. We summarize and expand on these suggestions.

When to use affordance augmentations? Our study findings suggest that visual augmentations should maximize utility and minimize visual clutter and confusion. As such, it is critical to render augmentations tailored to users’ intent and reduce distraction. For example, affordance augmentations that involve multiple pieces and colors are more preferred to support direct hand-object interactions, while whole object augmentations are more suitable in general visual perception tasks such as avoiding obstacles and locating an object. 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 Lang *et al.* [36]).

Where to apply affordance augmentations? Affordances can refer to any object parts that indicate diverse actions or interaction opportunities. However, LV people face distinct interaction challenges, resulting in unique affordance opportunities. In our qualitative study, we identified seven essential affordances of kitchen tools that encapsulate important yet challenging interaction 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 alignment 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 interaction between two objects; and (7) activation area, affording control features on an object. This affordance taxonomy 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 affordances according to the interaction tasks they indicate. In our study, participants preferred solid-colored overlays for grabbable areas 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 occlusion. In terms of colors, augmentations should adopt colors with high contrast against the environment. We also suggest leveraging cultural and semantic meanings of colors, such as green for safe-to-grab areas and red for risky areas. However, while preferred by the LV participants in our study, the green-red combination should be used cautiously given the prominence of red-green color vision deficiency.

How to control affordance augmentations? Due to the diverse visual abilities and preferences of LV users [81], future systems should support extensive personalization capabilities such as voice-based control for customization and automatic adaptations. For instance, users should be able to adjust different aspects of an augmentation, such as switching it on or off, choosing between solid-colored and outlined overlays, changing the outline thickness, and selecting suitable colors. Additionally, these systems should intelligently adapt by recognizing user actions or tasks to only highlight necessary object parts and signal warnings. They should also automatically alter the augmentation colors to generate high contrast against the background.

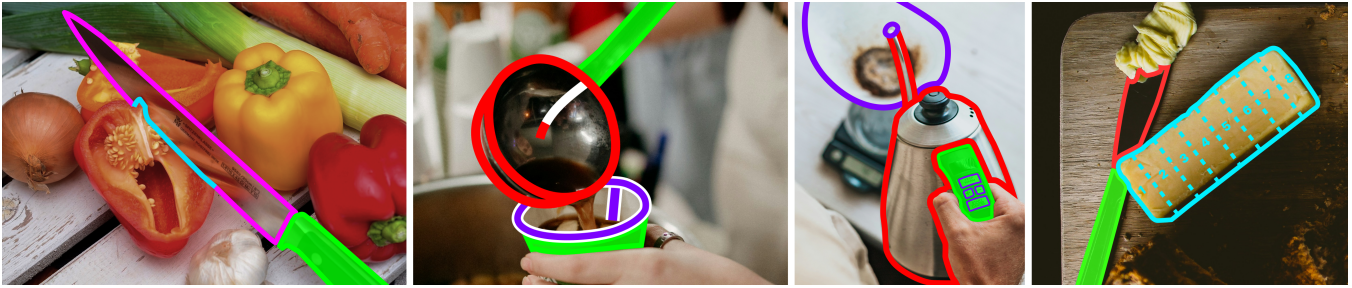


Figure 7: Envisioned augmentations based on participants’ desired designs, including (1) highlighting not just the grabbable and hazard areas of a knife (using a color pairing that is not green and red for increased contrast) but also the contact location with the target object (e.g., a bell pepper); (2) illuminating the exit and entry areas of a ladle and cup along with their containment areas and indicating the depth and filled-up level of the containment area; (3) showing the activation areas (i.e., buttons) on a carafe along with the exit location of the spout; (4) showing the intersection of a butter knife with butter as well as measurement highlights overlaid on a stick of butter (e.g., tablespoons).

Example designs. Reflecting on these key design insights for affordance augmentations, we created some initial design mockups shown in Figure 7. Starting with solid-colored overlays for grabbable areas, measuring lines for containment areas, and outlines for all other areas, CookAR should allow customization, such as toggling overlays, adjusting outline thickness, and changing colors. Future work should explore these and other designs empirically.

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. However, our study also revealed the impact of technology limitations on user experiences. For example, while finding CookAR promising, participants gave significantly higher Likert scores for the real-world baseline condition (i.e., without CookAR). 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, its mAP is still too low to successfully support dynamic activities like cooking in real-world contexts. 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, resulting in misalignment between augmentations and the original objects. The recognition results could become worse during real-world use on AR glasses due to natural human behaviors like users’ constant head motions. For example, LV users tend to get much closer to view objects than sighted users [64]. This issue highlights the evaluation gap between HCI and AI: a model that performs well under AI metrics (e.g., mAP, AP@50, AP@75) may be suitable for in-lab user studies but less so in naturalistic settings. 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.

Affordance models and datasets. As opposed to object recognition models and datasets that attract significant attention in

AI [13, 23, 40, 41], 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 speed. However, due to the relatively small scale of the dataset and the RTMDet model not being designed for affordance, our system encountered affordance-specific issues. For example, our model often struggled to distinguish different handles, as many handles across various cooking tools look similar. While not affecting the mask generation (allowing users to still see the correct augmentations), it interfered with the object tracking model supported by the ZED Mini API, leading to flickering and unstable augmentation rendering. This is also noticeable in the video supplement.

To enhance system robustness and affordance recognition capabilities, future AI research should consider the following key areas: (1) Developing larger and more diverse affordance datasets. These datasets should capture a wider variety of object interactions and functionalities, allowing the model to learn from a richer set of scenarios; (2) Designing models for affordance detection, such as incorporating training objectives that refine object part relationship understanding for better affordance prediction; and (3) Improving object tracking algorithms to ensure more stable augmentation rendering, especially in dynamic environments where precise object localization is crucial.

System latency. Latency is always a concern for AR systems, especially since off-the-shelf AR devices usually do not have sufficient computational power (e.g., GPU) to support real-time CV. To enable affordance segmentation, our system streams video data between the AR headset and an external server. However, system latency prevented the overlays from keeping pace with the participants’ head motions, negatively impacting their trust in CookAR’s intelligence and perceived system usability. To address system latency, we need advancements in both software (e.g., real-time AI models) and hardware (e.g., AR devices with powerful GPUs), which will also increase the overall usability of AI-powered AR systems in dynamic real-world activities.

7.3 Limitations & Future Directions

We outline four 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 AR-based affordance augmentations. Future work should conduct larger scale studies with more participants and diverse visual conditions. Second, we re-emphasize the aforementioned technical challenges and limitations. Third, 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 adjustments (e.g., colors, thickness of the outlines) to provide LV users more personalized experience. Finally, current CookAR system focuses on leveraging CV methods to detect affordances. While some suggested affordances, such as entrance, exit, and activation areas, can be achieved by dataset extension and model fine-tuning, others may not. For example, detecting heated areas may require a thermal sensor and identifying intersection areas can benefit from a LiDAR sensor. Future research should consider additional sensors beyond RGB cameras.

8 CONCLUSION

In this paper, we introduce CookAR, a wearable AR system that overlays 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 created an AR system with a stereo depth camera to generate real-time affordance augmentations in 3D space. We evaluated CookAR in a three-part lab study with 10 LV participants. Findings indicate participants' preferences for affordance augmentations over whole object augmentations for tool interactions, as well as revealing seven types of tool affordances and corresponding augmentation designs preferred by LV users. Our work highlights the promise of affordance augmentations in supporting hand-object interactions for LV people and advances state-of-the-art AI-powered AR technology as low vision aids.

ACKNOWLEDGMENTS

This work was supported by an NSF Graduate Research Fellowship and NSF CHS #1763199. We also thank Yang Li, Sieun Kim, and XunMei Liu for their help with our open-sourced content.

REFERENCES

- [1] Shiri Azenkot and Yuhang Zhao. 2017. Designing smartglasses applications for people with low vision. *SIGACCESS Access. Comput.* 119 (nov 2017), 19–24. <https://doi.org/10.1145/3167902.3167905>
- [2] Shikhar Bahl, Russell Mendonca, Lili Chen, Unnat Jain, and Deepak Pathak. 2023. Affordances from Human Videos as a Versatile Representation for Robotics. *CVPR*.
- [3] Marie Claire Bilyk, Jessica M Sontrop, Gwen E Chapman, Susan I Barr, and Linda Mamer. 2009. Food experiences and eating patterns of visually impaired and blind people. *Canadian Journal of Dietetic practice and research* 70, 1 (2009), 13–18.
- [4] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- [5] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- [6] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. 2019. MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv preprint arXiv:1906.07155* (2019).
- [7] Fu-Jen Chu, Ruinian Xu, Landan Seguin, and Patricio A. Vela. 2019. Toward Affordance Detection and Ranking on Novel Objects for Real-World Robotic Manipulation. *IEEE Robotics and Automation Letters* 4, 4 (2019), 4070–4077. <https://doi.org/10.1109/LRA.2019.2930364>
- [8] Fu-Jen Chu, Ruinian Xu, and Patricio A. Vela. 2019. Learning Affordance Segmentation for Real-World Robotic Manipulation via Synthetic Images. *IEEE Robotics and Automation Letters* 4, 2 (2019), 1140–1147. <https://doi.org/10.1109/LRA.2019.2894439>
- [9] Ching-Yao Chuang, Jiaman Li, Antonio Torralba, and Sanja Fidler. 2018. Learning to Act Properly: Predicting and Explaining Affordances from Images. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 975–983. <https://doi.org/10.1109/CVPR.2018.00108>
- [10] Anne Lesley Corn and Jane N Erin. 2010. *Foundations of low vision: Clinical and functional perspectives*. American Foundation for the Blind.
- [11] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, and Michael Wray. 2018. Scaling Egocentric Vision: The EPIC-KITCHENS Dataset. In *Proceedings of the European Conference on Computer Vision (ECCV)*.
- [12] Ashley D Deemer, Christopher K Bradley, Nicole C Ross, Danielle M Natale, Rath Itthiphanichpong, Frank S Werblin, and Robert W Massof. 2018. Low vision enhancement with head-mounted video display systems: are we there yet? *Optometry and Vision Science* 95, 9 (2018), 694–703.
- [13] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.
- [14] MR Everingham, BT Thomas, T Troscianko, et al. 1999. Head-mounted mobility aid for low vision using scene classification techniques. *The International Journal of Virtual Reality* 3, 4 (1999), 3.
- [15] Leah Findlater, Bonnie Chinh, Dhruv Jain, Jon Froehlich, Raja Kushalnagar, and Angela Carey Lin. 2019. Deaf and Hard-of-hearing Individuals' Preferences for Wearable and Mobile Sound Awareness Technologies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300276>
- [16] Dylan R. Fox, Ahmad Ahmadzada, Clara Tenia Wang, Shiri Azenkot, Marlena A. Chu, Roberto Manduchi, and Emily A. Cooper. 2023. Using augmented reality to cue obstacles for people with low vision. *Opt. Express* 31, 4 (Feb 2023), 6827–6848. <https://doi.org/10.1364/OE.479258>
- [17] Patricia Garrido-Vásquez and Anna Schubö. 2014. Modulation of visual attention by object affordance. *Frontiers in Psychology* 5 (2014), 70664.
- [18] James Jerome Gibson. 1966. The senses considered as perceptual systems. (1966).
- [19] James J Gibson. 1977. The theory of affordances. *Hilldale, USA* 1, 2 (1977), 67–82.
- [20] James J Gibson. 2014. *The ecological approach to visual perception: classic edition*. Psychology press.
- [21] Anshul Gupta, Juraj Mesik, Stephen A Engel, Rebecca Smith, Mark Schatza, Aurélie Calabrese, Frederik J Van Kuijk, Arthur G Erdman, and Gordon E Legge. 2018. Beneficial effects of spatial remapping for reading with simulated central field loss. *Investigative ophthalmology & visual science* 59, 2 (2018), 1105–1112.
- [22] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*. 2961–2969.
- [23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [24] Stephen L Hicks, Iain Wilson, Louwai Muhammed, John Worsfold, Susan M Downes, and Christopher Kennard. 2013. A depth-based head-mounted visual display to aid navigation in partially sighted individuals. *PLoS one* 8, 7 (2013), e67695.
- [25] Jonathan Huang, Max Kineteder, Matt J Dunn, Wojciech Jarosz, Xing-Dong Yang, and Emily A Cooper. 2019. An augmented reality sign-reading assistant for users with reduced vision. *PLoS one* 14, 1 (2019), e0210630.
- [26] Alex D Hwang and Eli Peli. 2014. An augmented-reality edge enhancement application for Google Glass. *Optometry and vision science* 91, 8 (2014), 1021–1030.
- [27] Dhruv Jain, Bonnie Chinh, Leah Findlater, Raja Kushalnagar, and Jon Froehlich. 2018. Exploring Augmented Reality Approaches to Real-Time Captioning: A Preliminary Autoethnographic Study. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems* (Hong Kong, China) (DIS '18 Companion). Association for Computing Machinery, New York, NY, USA, 7–11. <https://doi.org/10.1145/3197391.3205404>

- [28] Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E. Froehlich. 2015. Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (CHI '15). Association for Computing Machinery, New York, NY, USA, 241–250. <https://doi.org/10.1145/2702123.2702393>
- [29] Dhruv Jain, Rachel Franz, Leah Findlater, Jackson Cannon, Raja Kushalnagar, and Jon Froehlich. 2018. Towards Accessible Conversations in a Mobile Context for People who are Deaf and Hard of Hearing. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) (ASSETS '18). Association for Computing Machinery, New York, NY, USA, 81–92. <https://doi.org/10.1145/3234695.3236362>
- [30] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. 2023. *Ultralytics YOLO*. <https://github.com/ultralytics/ultralytics>
- [31] Nabila Jones, Hannah Elizabeth Bartlett, and Richard Cooke. 2019. An analysis of the impact of visual impairment on activities of daily living and vision-related quality of life in a visually impaired adult population. *British Journal of Visual Impairment* 37, 1 (2019), 50–63.
- [32] Avyay Ravi Kashyap. 2020. Behaviors, Problems and strategies of visually impaired persons during meal preparation in the Indian context: challenges and opportunities for Design. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–3.
- [33] Minyung Kim, Sooyoung Hwang, Kyoungmin Choi, Youkeun Oh, and Dokshin Lim. 2022. Vision-Based Cooking Assistance System for Visually Impaired People. In *International Conference on Human-Computer Interaction*. Springer, 540–547.
- [34] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. *arXiv:2304.02643* (2023).
- [35] MiYoung Kwon, Chaithanya Ramachandra, PremNandhini Satgunam, Bartlett W Mel, Eli Peli, and Bosco S Tjan. 2012. Contour enhancement benefits older adults with simulated central field loss. *Optometry and vision science* 89, 9 (2012), 1374–1384.
- [36] Florian Lang and Tonja Machulla. 2021. Pressing a Button You Cannot See: Evaluating Visual Designs to Assist Persons with Low Vision through Augmented Reality. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (Osaka, Japan) (VRST '21). Association for Computing Machinery, New York, NY, USA, Article 39, 10 pages. <https://doi.org/10.1145/3489849.3489873>
- [37] Jaewook Lee, Devesh P. Sarda, Eujean Lee, Amy Lee, Jun Wang, Adrian Rodriguez, and Jon E. Froehlich. 2023. Towards Real-time Computer Vision and Augmented Reality to Support Low Vision Sports: A Demonstration of ARTennis. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (, San Francisco, CA, USA.) (UIST '23 Adjunct). Association for Computing Machinery, New York, NY, USA, Article 81, 3 pages. <https://doi.org/10.1145/3586182.3615815>
- [38] Franklin Mingzhe Li, Jamie Dorst, Peter Cederberg, and Patrick Carrington. 2021. Non-Visual Cooking: Exploring Practices and Challenges of Meal Preparation by People with Visual Impairments. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (, Virtual Event, USA.) (ASSETS '21). Association for Computing Machinery, New York, NY, USA, Article 30, 11 pages. <https://doi.org/10.1145/3441852.3471215>
- [39] Franklin Mingzhe Li, Michael Xieyang Liu, Shaun K Kane, and Patrick Carrington. 2024. A Contextual Inquiry of People with Vision Impairments in Cooking. *arXiv preprint arXiv:2402.15108* (2024).
- [40] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*. Springer, 740–755.
- [41] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*. 10012–10022.
- [42] David S Loshin and Richard D Juday. 1989. The programmable remapper: clinical applications for patients with field defects. *Optometry and Vision Science* 66, 6 (1989), 389–395.
- [43] Timo Luddecke and Florentin Worgotter. 2017. Learning to segment affordances. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 769–776.
- [44] Gang Luo and Eli Peli. 2006. Use of an augmented-vision device for visual search by patients with tunnel vision. *Investigative ophthalmology & visual science* 47, 9 (2006), 4152–4159.
- [45] Hongchen Luo, Wei Zhai, Jing Zhang, Yang Cao, and Dacheng Tao. 2022. Learning Affordance Grounding from Exocentric Images. In *CVPR*.
- [46] Chengqi Lyu, Wenwei Zhang, Haiyan Huang, Yue Zhou, Yudong Wang, Yanyi Liu, Shilong Zhang, and Kai Chen. 2022. RTMDet: An Empirical Study of Designing Real-Time Object Detectors. *arXiv:2212.07784* [cs.CV]
- [47] Meethu Malu and Leah Findlater. 2014. "OK Glass?" A Preliminary Exploration of Google Glass for Persons with Upper Body Motor Impairments. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility* (Rochester, New York, USA) (ASSETS '14). Association for Computing Machinery, New York, NY, USA, 267–268. <https://doi.org/10.1145/2661334.2661400>
- [48] Meethu Malu and Leah Findlater. 2015. Personalized, Wearable Control of a Head-mounted Display for Users with Upper Body Motor Impairments. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (CHI '15). Association for Computing Machinery, New York, NY, USA, 221–230. <https://doi.org/10.1145/2702123.2702188>
- [49] Robert W Massof and Douglas L Rickman. 1992. Obstacles encountered in the development of the low vision enhancement system. *Optometry and vision science* 69, 1 (1992), 32–41.
- [50] Robert W Massof, Douglas L Rickman, Peter A Lalle, et al. 1994. Low vision enhancement system. *Johns Hopkins APL Technical Digest* 15, 2 (1994), 120–125.
- [51] Roisin McNaney, John Vines, Daniel Roggen, Madeline Balaam, Pengfei Zhang, Ivan Poliakov, and Patrick Olivier. 2014. Exploring the acceptability of google glass as an everyday assistive device for people with parkinson's. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (CHI '14). Association for Computing Machinery, New York, NY, USA, 2551–2554. <https://doi.org/10.1145/2556288.2557092>
- [52] Ashley Miller, Joan Malasig, Brenda Castro, Vicki L. Hanson, Hugo Nicolau, and Alessandra Brandão. 2017. The Use of Smart Glasses for Lecture Comprehension by Deaf and Hard of Hearing Students. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (, Denver, Colorado, USA.) (CHI EA '17). Association for Computing Machinery, New York, NY, USA, 1909–1915. <https://doi.org/10.1145/3027063.3053117>
- [53] Susanna Mills, Martin White, Heather Brown, Wendy Wrieden, Dominika Kwasnicka, Joel Halligan, Shannon Robalino, and Jean Adams. 2017. Health and social determinants and outcomes of home cooking: A systematic review of observational studies. *Appetite* 111 (2017), 116–134.
- [54] Kenzaburo Miyawaki, Mutsuo Sano, Syunichi Yonemura, and Michiko Ode. 2012. Cooking rehabilitation support for self-reliance of cognitive dysfunction patients. In *Proceedings of the ACM multimedia 2012 workshop on Multimedia for cooking and eating activities*. 19–24.
- [55] Pilar Montero. 2005. Nutritional assessment and diet quality of visually impaired Spanish children. *Annals of Human Biology* 32, 4 (2005), 498–512.
- [56] Elizabeth D Mynatt and Wendy A Rogers. 2001. Developing technology to support the functional independence of older adults. *Ageing International* 27, 1 (2001), 24–41.
- [57] Anh Nguyen, Dimitrios Kanoulas, Darwin G. Caldwell, and Nikos G. Tsagarakis. 2016. Detecting object affordances with Convolutional Neural Networks. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2765–2770. <https://doi.org/10.1109/IROS.2016.7759429>
- [58] Anh Nguyen, Dimitrios Kanoulas, Darwin G. Caldwell, and Nikos G. Tsagarakis. 2017. Object-based affordances detection with Convolutional Neural Networks and dense Conditional Random Fields. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 5908–5915. <https://doi.org/10.1109/IROS.2017.8206484>
- [59] Alex Olwal, Kevin Balke, Dmitrii Votintsev, Thad Starner, Paula Conn, Bonnie Chinh, and Benoit Corda. 2020. Wearable Subtitles: Augmenting Spoken Communication with Lightweight Eyewear for All-day Captioning. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) (UIST '20). Association for Computing Machinery, New York, NY, USA, 1108–1120. <https://doi.org/10.1145/3379337.3415817>
- [60] Eli Peli. 2001. Vision multiplexing: an engineering approach to vision rehabilitation device development. *Optometry and Vision Science* 78, 5 (2001), 304–315.
- [61] Yi-Hao Peng, Ming-Wei Hsi, Paul Tael, Ting-Yu Lin, Po-En Lai, Leon Hsu, Tzu-chuan Chen, Te-Yen Wu, Yu-An Chen, Hsien-Hui Tang, and Mike Y. Chen. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (, Montreal QC, Canada.) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–10. <https://doi.org/10.1145/3173574.3173867>
- [62] Kyle Rector, Lauren Milne, Richard E Ladner, Batya Friedman, and Julie A Kientz. 2015. Exploring the opportunities and challenges with exercise technologies for people who are blind or low-vision. In *Proceedings of the 17th international ACM SIGACCESS conference on computers & accessibility*. 203–214.
- [63] Gwendolyn Rehrig, Madison Barker, Candace E Peacock, Taylor R Hayes, John M Henderson, and Fernanda Ferreira. 2022. Look at what I can do: Object affordances guide visual attention while speakers describe potential actions. *Attention, Perception, & Psychophysics* 84, 5 (2022), 1583–1610.
- [64] Jarek Reynolds, Chandra Kanth Nagesh, and Danna Gurari. 2024. Salient object detection for images taken by people with vision impairments. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 8522–8531.

- [65] Chris Schipper and Bo Brinkman. 2017. Caption Placement on an Augmented Reality Head Worn Device. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) (ASSETS '17). Association for Computing Machinery, New York, NY, USA, 365–366. <https://doi.org/10.1145/3132525.3134786>
- [66] Lee Stearns, Victor DeSouza, Jessica Yin, Leah Findlater, and Jon E. Froehlich. 2017. Augmented Reality Magnification for Low Vision Users with the Microsoft HoloLens and a Finger-Worn Camera. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) (ASSETS '17). Association for Computing Machinery, New York, NY, USA, 361–362. <https://doi.org/10.1145/3132525.3134812>
- [67] Stereolabs. 2024. zed-unity. <https://github.com/stereolabs/zed-unity>.
- [68] Sarit Szpiro, Yuhang Zhao, and Shiri Azenkot. 2016. Finding a store, searching for a product: a study of daily challenges of low vision people. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) (UbiComp '16). Association for Computing Machinery, New York, NY, USA, 61–72. <https://doi.org/10.1145/2971648.2971723>
- [69] Sarit Felicia Anais Szpiro, Shafeka Hashash, Yuhang Zhao, and Shiri Azenkot. 2016. How People with Low Vision Access Computing Devices: Understanding Challenges and Opportunities. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (Reno, Nevada, USA) (ASSETS '16). Association for Computing Machinery, New York, NY, USA, 171–180. <https://doi.org/10.1145/2982142.2982168>
- [70] Deanna J Taylor, Angharad E Hobby, Alison M Binns, and David P Crabb. 2016. How does age-related macular degeneration affect real-world visual ability and quality of life? A systematic review. *BMJ open* 6, 12 (2016), e011504.
- [71] Jan Tünnermann, Norbert Krüger, Bärbel Mertsching, and Wail Mustafa. 2015. Affordance estimation enhances artificial visual attention: Evidence from a change-blindness study. *Cognitive computation* 7 (2015), 526–538.
- [72] Joram J van Rheede, Iain R Wilson, Rose I Qian, Susan M Downes, Christopher Kennard, and Stephen L Hicks. 2015. Improving mobility performance in low vision with a distance-based representation of the visual scene. *Investigative ophthalmology & visual science* 56, 8 (2015), 4802–4809.
- [73] Ru Wang, Nihan Zhou, Tam Nguyen, Sanbrita Mondal, Bilge Mutlu, and Yuhang Zhao. 2023. Characterizing Barriers and Technology Needs in the Kitchen for Blind and Low Vision People. *arXiv preprint arXiv:2310.05396* (2023).
- [74] Ru Wang, Nihan Zhou, Tam Nguyen, Sanbrita Mondal, Bilge Mutlu, and Yuhang Zhao. 2023. Practices and Barriers of Cooking Training for Blind and Low Vision People. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (, New York, NY, USA,) (ASSETS '23). Association for Computing Machinery, New York, NY, USA, Article 57, 5 pages. <https://doi.org/10.1145/3597638.3614494>
- [75] Peter Washington, Catalin Voss, Aaron Kline, Nick Haber, Jena Daniels, Azar Fazel, Titus De, Carl Feinstein, Terry Winograd, and Dennis Wall. 2017. SuperpowerGlass: A Wearable Aid for the At-Home Therapy of Children with Autism. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 112 (sep 2017), 22 pages. <https://doi.org/10.1145/3130977>
- [76] Kristin Williams, Karyn Moffatt, Denise McCall, and Leah Findlater. 2015. Designing Conversation Cues on a Head-Worn Display to Support Persons with Aphasia. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (CHI '15). Association for Computing Machinery, New York, NY, USA, 231–240. <https://doi.org/10.1145/2702123.2702484>
- [77] Yuhang Zhao, Michele Hu, Shafeka Hashash, and Shiri Azenkot. 2017. Understanding Low Vision People's Visual Perception on Commercial Augmented Reality Glasses. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 4170–4181. <https://doi.org/10.1145/3025453.3025949>
- [78] Yuhang Zhao, Elizabeth Kupferstein, Brenda Veronica Castro, Steven Feiner, and Shiri Azenkot. 2019. Designing AR Visualizations to Facilitate Stair Navigation for People with Low Vision. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) (UIST '19). Association for Computing Machinery, New York, NY, USA, 387–402. <https://doi.org/10.1145/3332165.3347906>
- [79] Yuhang Zhao, Elizabeth Kupferstein, Hathaithorn Rojnirun, Leah Findlater, and Shiri Azenkot. 2020. The Effectiveness of Visual and Audio Wayfinding Guidance on Smartglasses for People with Low Vision. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (, Honolulu, HI, USA,) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376516>
- [80] Yuhang Zhao, Elizabeth Kupferstein, Doron Tal, and Shiri Azenkot. 2018. "It Looks Beautiful but Scary" How Low Vision People Navigate Stairs and Other Surface Level Changes. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 307–320.
- [81] Yuhang Zhao, Sarit Szpiro, and Shiri Azenkot. 2015. Foresee: A customizable head-mounted vision enhancement system for people with low vision. In *Proceedings of the 17th international ACM SIGACCESS conference on computers & accessibility*. 239–249.
- [82] Yuhang Zhao, Sarit Szpiro, Jonathan Knighten, and Shiri Azenkot. 2016. CueSee: exploring visual cues for people with low vision to facilitate a visual search task. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) (UbiComp '16). Association for Computing Machinery, New York, NY, USA, 73–84. <https://doi.org/10.1145/2971648.2971730>

A PART 1 PROTOCOL

In Part 1, participants compared real-world baseline (*i.e.*, using their typical method in daily life), augmentation baseline (*i.e.*, whole object augmentations), and CookAR (*i.e.*, affordance augmentations) in a tool-grabbing task. We asked participants Likert questions and open-ended questions after each condition. After Part 1, we asked additional qualitative questions.

A.1 7-Point Likert Scale Questionnaire

1. How effective is the system you just used? Why?
2. How comfortable are you with seeing these visualizations? Why?
3. How distracting are these visualizations? Why?

A.2 Post-Condition Open-Ended Questions

1. Please describe your overall experience with the visual augmentations.
2. What did you like about this system?
3. What did you dislike about this system?
4. (If any) Ask participants about any interactions we observed to be unconventional (*e.g.*, grabbing a knife by the blade)
5. How can this system be improved?
6. In what other scenarios do you think this system would be useful?
7. Do you have any additional comments about this system that we failed to capture?

A.3 Post-Part Open-Ended Questions

1. Can you compare your experience completing this study task with and without our research prototypes?
2. Which condition do you prefer to use? Why?
3. How can this system be improved? Are there any design recommendations you'd like to make?
4. In which scenarios do you think this system would be useful?
5. Do you have any additional comments about this study or the systems?

B PART 2 PROTOCOL

In Part 2, participants completed a full cooking task using CookAR. They then answered open-ended questions about their experience.

1. Please describe your overall experience cooking with this system.
2. What did you like about this system?
3. What did you dislike about this system?
4. What are some improvements you want to make to the system?
5. Are there any design recommendations you would like to make?
6. In which scenarios do you think this system would be useful?
7. Any additional comments about this study or this system?

C PART 3 PROTOCOL

In Part 3, participants brainstormed future designs and applications of CookAR. We asked open-ended questions to guide them.

1. Can you think of your own design that would improve the usability of our prototype?
2. In what other scenarios besides cooking do you think this kind of a system can be applicable?
3. Any additional comments about the system and the entire user study?