Improving Context-Aware Habit-Support Interventions Using Egocentric Visual Contexts

Mina Khan[1(✉)], Glenn Fernandes[1], Akash Vaish[1], Mayank Manuja[1], Pattie Maes[1], and Agnis Stibe[2,3]

[1] MIT Media Lab, 75 Amherst Street, Cambridge, MA, USA
glennfer@mit.edu, pattie@media.mit.edu
[2] Metis Lab, EM Normandie Business School, Le Havre, France
agnis@transforms.me
[3] INTERACT Research Unit, University of Oulu, Oulu, Finland

Abstract. Habits are automatic actions in stable contexts and can help sustain behavior change. State-of-the-art context-aware habit-support interventions are, however, in predominantly non-visual contexts (time, geolocation, and physical activity), and not in egocentric visual contexts, e.g., eating, meeting people, etc. Using a survey, N = 51 participants, we identified the user-desired visual contexts and interventions mediums for habit support. Based on our survey, we created a wearable system, named PAL, with privacy-preserving on-device deep learning, to deliver real-world habit-support interventions in personalized visual contexts. In our 4-week study, N = 10 participants, interventions using PAL's personalized visual contexts led to >75% increase in habit formation, compared to <40% increase in habit formation using interventions in only non-visual contexts. The habits also persisted in the post-study evaluations 1 and 10 weeks later. Thus, PAL's interventions using personalized visual contexts improve real-world habit formation for sustainable behavior change.

Keywords: Persuasive technology · Wearable · Context-aware · Interventions · Habits · Personalized · Deep learning · Visual contexts

1 Introduction

Persuasive Technologies aim to support behavior change, but commonly used behavior change techniques like reminders induce dependency and behavior change does not persist after the users stop using the apps [31]. Habits are automatic actions in stable contexts [28] and can help sustain behavior change [17,40].

Triggers/contexts have been key for Persuasive Technologies [6] and context-aware technologies can provide interventions in automatically-detected contexts in our everyday lives. State-of-the-art context-aware habit-support interventions

R. Ali et al. (Eds.): PERSUASIVE 2021, LNCS 12684, pp. 115–131, 2021.
https://doi.org/10.1007/978-3-030-79460-6_10

are in contexts involving geolocation, physical motion, and time [29]. However, research shows that users want habit-support interventions in more contexts, e.g., indoor locations and specific objects [29]. Also, habit-formation research has used contexts like brushing teeth [13] and lunch [34], which are currently not automatically detected to deliver context-aware habit-support interventions.

Deep learning-based computer vision models can detect diverse visual contexts, e.g., objects, [37] to provide more information about the user's context. We explore if context-aware interventions using egocentric visual contexts, e.g., eating, brushing teeth, etc., can improve real-world habit-support interventions.

We conducted a user survey with 51 participants (Sect. 3) to identify the desired habit-support intervention contexts and mediums. Our survey showed that people want habit-support in five types of visual contexts (generic faces, objects, custom faces, custom activities, and custom indoor locations), but are wary of using wearable cameras because of privacy concerns. In addition, audio output was the most desired intervention medium for habit support.

In light of our survey, we created a wearable system, named PAL (Sect. 4), to deliver open-ear audio interventions in personalized visual contexts for habit support. Considering camera-related privacy concerns, we used on-device deep learning so that user data is not sent to the cloud/another device for model training or inference. We used deep learning models, tested in real-life settings, to recognize the five types of visual contexts highlighted in our survey.

We designed a 4-week habit-formation study (Sect. 5) with 10 participants to compare the efficacy of PAL's habit-support interventions using personalized visual contexts with interventions using non-visual contexts. Our results (Sect. 6) show more habit formation with interventions using visual contexts (>75% increase) than with interventions using only non-visual contexts, i.e., geolocation, physical activity, and time (<40% increase). The habits also persisted in the post-study surveys 1 and 10 weeks later. We discuss our findings in Sect. 7.

We make three contributions: i. a survey, N = 51 participants, identifying the user-desired visual contexts and intervention mediums for habit support; ii. a wearable system, named PAL, with privacy-preserving on-device deep learning, to deliver habit-support interventions in personalized visual contexts; iii. a 4-week study, N = 10 participants, with 1 and 10 week later post-study evaluations, showing almost double habit formation with PAL's interventions using personalized visual contexts than with interventions using only non-visual contexts.

2 Related Work

Contexts have always been key to behavior change and persuasive technologies. The Fogg Behavior Model [6] recommends triggers, which tie new behaviors to existing contexts/routines. The need for context-awareness and just-in-time interventions has also been highlighted for persuasive technologies [12,35].

Our work leverages interdisciplinary insights from habit formation and deep learning to create a wearable system for just-in-time habit-support interventions

using personalized visual context detection. Our related work falls into three categories: i. Context-based Habit-support Interventions, ii. Context-aware (Non-habit) Behavior Change Interventions, and iii. Wearable Visual Context Detection. To the best of our knowledge, there are no wearable systems for just-in-time habit-support interventions using personalized visual context detection.

2.1 Context-Based Habit-Support Interventions

Instead of time-based reminders, research suggests leveraging the context-based nature of habits to avoid forgetfulness, e.g., by tying medication reminders to existing routines [33]. Time-based reminders have been shown to have higher adherence but lower automaticity than event-based triggers (without reminders), e.g., after lunch [34]. Researchers have used 'plan reminders' to remind the users of their context-based habit goals [38,39], but these are not in automatically recognized contexts. Non-visual contexts, i.e., time, location, and physical activity, have been used for habit-support interventions [29] in automatically detected contexts, but there are no automatic context-aware habit-support interventions in personalized visual contexts.

2.2 Context-Aware (Non-habit) Behavior Change Interventions

Context-aware interventions, not focused on habit formation, have been investigated in context-aware behavior change systems [9], e.g., using location [26], computer usage [26], physiological signals [2,19,25], multimodal sensing (e.g., heart rate, movement, and computer usage [14]), and even locally-installed motion sensors for activity sensing [22]. Unlike the existing non-visual context detection techniques, wearable visual context detection can recognize faces, objects, activities, scenes, etc., in a mobile context. However, there are no wearable systems for just-in-time behavior change interventions using personalized visual contexts.

2.3 Wearable Visual Context Detection

Deep learning-based egocentric visual context recognition exists, e.g., for memory support [20,21] and visual assistance [1,27], and some systems even distort images for enhancing privacy [4]. There are also on-device deep learning systems for computer vision [18,24]. However, unlike PAL, there are no wearable systems with on-device deep learning for privacy-preserving detection of personalized visual contexts for context-aware habit or behavior change support.

3 Habit-Support Interventions Design

In order to identify the user preferences for habit-support interventions in visual contexts, we conducted a survey about the desired habit-support intervention contexts, intervention mediums, and context detection preferences: *"Think of a habit you would like to develop, i.e. what and when would you do. Q1. When*

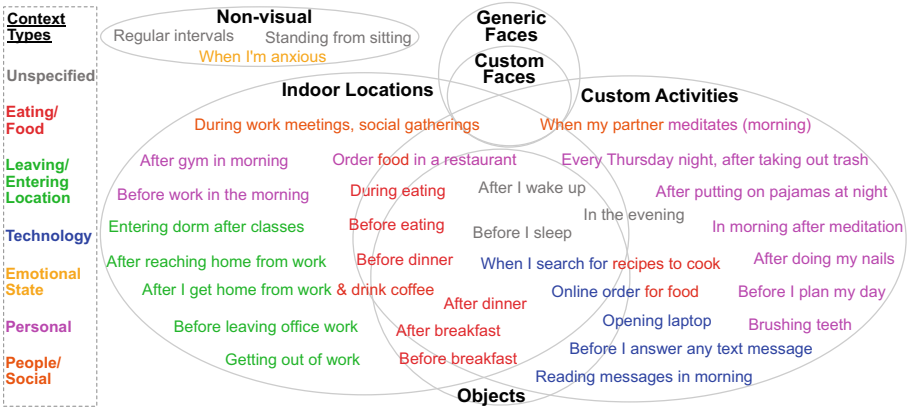

Fig. 1. Open-ended survey responses, N = 51 participants, for the desired habit-support intervention contexts, grouped by one non-visual and five possibly visual categories.

would you like interventions? (Intervention Contexts); Q2. How would you like the intervention? (Intervention Mediums); Q3. Would you like interventions in visual contexts? (no/maybe/yes); Q4. Would you like to use a camera? (no/maybe/yes); Q5. Why/why not?"

We recruited 51 participants ($\mu = 29$ yrs, $\sigma = 10.85$ yrs; 20 males, 30 females; 5 countries; 14 students, 24 professionals, 13 unknown), without any inclusion or exclusion criteria, using our social media and department emailing list.

We summarize the survey results and corresponding design decisions below.

3.1 Intervention Contexts

Only 4 of the 51 desired intervention contexts (Q1) were strictly non-visual, i.e., involving only time, geolocation, or emotional state, while the rest involved visual contexts. We categorized the visual contexts using a combination of 5 broad categories, i.e., generic faces, generic objects, custom faces, custom activities, and custom indoor locations (Fig. 1). We decided to detect the five types of visual contexts for habit-support interventions, and allow users to choose a combination of contexts for habit-support interventions since some contexts involved multiple contexts, e.g., "meditation in morning" involves time and visual contexts.

3.2 Intervention Mediums

We categorized the open-ended responses (Q2). Audio output was the most popular (45%), followed by text notifications (25%), ambient notes (14%), text or audio messages (8%), no reminders (4%), and unknown (4%). Since audio interventions were the most desired, we chose wearable open-ear audio output to privately, seamlessly, and unobtrusively deliver interventions anywhere.

3.3 Cameras and Privacy

Many participants indicated (N = No, M = Maybe, Y = Yes) that they wanted "Interventions in visual contexts" (Q3: 2N, 16M, 33Y) but due to "privacy concerns" (open-ended responses, Q5), did not want to use "Wearable cameras" (Q4: 23N, 15M, 13Y). Thus, we decided to use on-device deep learning for visual context detection so that user images are not sent to the cloud/another device and can be automatically deleted after on-device model training and/or inference. Also, any images saved for user labeling of custom faces, contexts, and indoor locations are deleted right after labeling.

4 PAL Implementation

We developed a wearable system, named PAL, for context-aware habit-support interventions in egocentric visual contexts. PAL has a wearable device, with on-device deep learning, for interventions in personalized visual contexts, and a mobile app for goal-setting, data labeling, and non-visual context detection.

4.1 Mobile App

The mobile app supports goal-setting and intervention context selection Fig. 2a and b), data labeling (Fig. 2c), and non-visual context detection.

We used implementation intentions [8] for goal-setting and intervention context selection. Implementation intentions are "if-then" action plans, e.g., if 'leaving home', then 'pick up fruits', and are commonly used for setting habit goals and interventions contexts [29,38,39]. We allow a combination of two contexts (using AND/OR) for habit-support interventions and include a 30-minute interval between interventions to avoid too many interventions.

The mobile app is connected to the wearable device over Bluetooth to configure the wearable device (e.g., turn off the camera, set intervention contexts, etc.), and for accessing the phone's non-visual context data, i.e., geolocation and physical activity, collected via Google's Places and Activity Transition APIs.

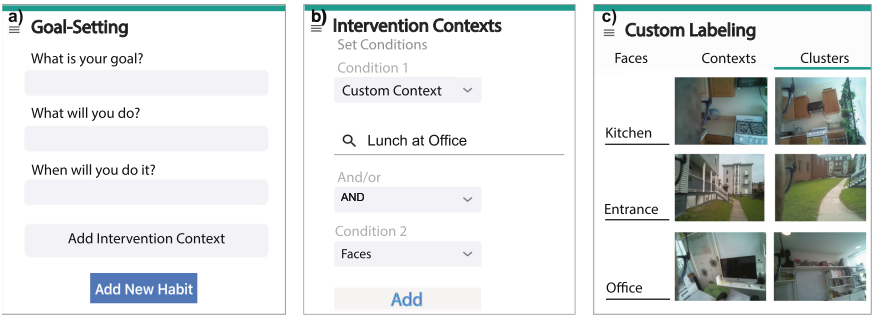

Fig. 2. Mobile app: (a) Goal-setting; (b) Context selection; (c) Custom labeling.

4.2 Wearable Device

PAL's wearable device has an on-body component connected to an on-ear component (Fig. 3). The on-ear component has a camera and speaker, and the on-body component has a microprocessor and a deep learning accelerator chip (Google Coral). There is also a button for taking custom training images.

We considered the on-ear placement suitable for not just open-ear audio output, but also the wearable camera. Cameras on non-face body parts are common, but due to their distance from the eyes, do not always capture the same scene as a user's eyes, especially when the user turns or tilts their head. Glasses are commonly used for on-face cameras but we decided to not use glasses frames as they are relatively bulky and pronounced on the face. Our on-ear camera captures ∼70% of a person's visual context (1200 cm × 750 cm view ∼1 m away).

On-device deep learning enables privacy-preserving context detection as the user images are not sent to the cloud/another device for training or processing. Also, it avoids time-consuming, power-hungry, and connectivity-dependent constant communication with the cloud/another device for real-time processing. The device consumes maximum 0.3A and our 2500 mAh battery lasts ∼5 h.

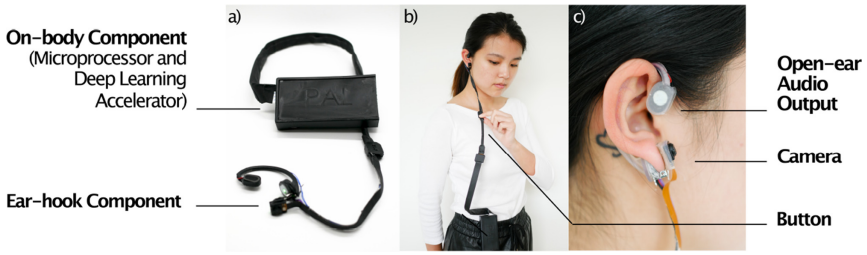

Fig. 3. (a) PAL's full wearable device with an on-device deep learning accelerator, (b) worn by a person, (c) with a close-up of the on-ear camera and audio output.

4.3 Context Detection Models

PAL has three types of models to recognize the aforementioned five types of visual contexts: i. Fully trained models for (generic) Object and Face Detection; ii. Low-shot custom-trainable models for Custom Face Recognition (1–2 training images) and Custom Context Recognition (for custom activities, ∼10 training images); iii. Semi-supervised model, i.e., Custom Context Clustering (for indoor locations separated by geolocations). The model details are in Table 1.

We chose 90-item Common Objects in Context (COCO) [23] dataset for object detection as COCO includes several common objects like a book, cup, toothbrush, etc. We chose Weight Imprinting [30] for the Custom Context Recognition model because it adds new classes to the old ones, instead of replacing

Table 1. Models training and architecture details for visual context recognition.

Visual contexts	Models
Object Detection	MobileNet SSD v2 [11] trained on 90-item COCO [23]
Faces Detection	MobileNet SSD v2 [11] trained on Open Images v4 [16]
Custom Faces Recognition	FaceNet [32], 1–2 training images/face
Custom Context Recognition	Weight imprinting [30] (MobileNet v1 pre-trained on 1000-class ImageNet [3], ∼10 training images/context
Custom Context Clustering	Image Embedding (MobileNet v1 pre-trained on 1000-class ImageNet) clustered via Density-Based Spatial Clustering of Applications with Noise [5]

Table 2. In-the-wild evaluations, N = ∼1000 images, of visual context models. (We list F1-score, in addition to the accuracy, for models tested with imbalanced classes).

Visual contexts	Results
Object Detection	98.8% accuracy, F1-score = 0.79 (∼1000 instances)
Faces Detection	88.8% accuracy, F1-score = 0.9 (∼180 faces)
Custom Faces Recognition	86.9% accuracy (4 known faces, 120 instances)
Custom Context Recognition	87.2% accuracy (7 contexts, ∼350 images)
Custom Context Clustering	82% accuracy (19 indoor locations, ∼300 images)

the old classes, and thus, the users can incrementally add more custom contexts over time. Custom Context Recognition is intended for visually similar contexts, e.g., activities like brushing teeth, whereas Custom Context Clustering is for clustering different, yet connected, views of a context, e.g., indoor locations.

All models are trained and inferred on the wearable device. The user presses a button on the device to start and stop a custom training session (6 images per minute) and labels the images on the mobile app. We tested the models with ∼1000 in-the-wild images of 4 users for 2 days (1 image every 2 min). Each model had a ≥70% accuracy and ∼3 s inference time. The results are in Table 2.

5 Study Design

We designed a study to compare context-aware habit-support interventions using only non-visual contexts, i.e., time, physical activity, and location, [29] *(Group Control)* with those using personalized visual context detection *(Group PAL)*.

5.1 Participants

We recruited 10 participants via our department email list (N = 10; $\mu = 23$ yrs, $\sigma = 2.36$ yrs; 7 males, 3 females; all students). We randomly created 2 groups -

Group PAL (N = 5; $\mu = 23.2$ yrs, $\sigma = 2.39$ yrs; 3 males, 2 females), and Group Control (N = 5; $\mu = 22.8$ yrs, $\sigma = 2.59$ yrs; 4 males, 1 female). Both groups used PAL's system, but only Group PAL had access to visual contexts for interventions.

5.2 Measures

We used three measures: *i. Weekly Habit-Formation Questionnaire*, *ii. Weekly Experience Questionnaire*, and *iii. End-of-study Open-ended Interview.*

Weekly Habit-Formation Questionnaire: Behavior change is an intricate and long-term process, and instead of measuring behavior change, it is recommended to do "efficacy evaluations", which are "tailored to the specific behavior change interventions" [15]. Since our interventions were aimed at habit-formation, we used habit-formation as an "efficacy measure" [15]. We used the Self-Report Habit Index (SRHI) [36] and Self-Report Behavioural Automaticity Index (SRBAI) [7] as they are commonly used to quantify habit formation [29,34,39].

Weekly Experience Questionnaire and *End-of-study Open-ended Interview*: It is suggested that it is too limiting to treat behavior change as a binary variable [15], and that research must help better understand the behavior change process [10]. In order to evaluate the habit-formation experiences of each participant, we sent a Weekly Experience Questionnaire and conducted an end-of-study in-person interview. The Weekly Experience Questionnaire had 3 open-ended questions – "How was your behavior change experience?", "Did the system help or hinder you? How/why?", and "Is there anything else you'd like to add?".

5.3 Procedures

We conducted a 4-week study, with post-study evaluations 1 and 10 weeks later, to monitor if the habits persisted after the study.

Free-living behavior change evaluations are recommended [10] and we attempted to keep our study as "free-living" [10] as possible. The participants could use the system whenever they wanted to but did not have to. The participants also did not get any financial compensation or other incentives for study completion or habit execution. Similar to Pinder et al.'s habit-formation study with interventions in only non-visual contexts [29], we allowed participants to select personalized habit goals and intervention contexts. We did not collect user's sensor data, e.g., images and geolocation, for privacy reasons.

At the start of the study, we explained the habit-support system to the participants, guiding them about how they can set a target habit and custom intervention context, including training personalized visual contexts. For their target habit, each participant filled the Weekly Habit-Formation Questionnaire at the beginning of the study, at the end of every week for the 4-week study, and also 1 and 10 weeks after the study for post-study evaluations. At the end of every week during the study, the participants also filled the aforementioned

Table 3. Intervention contexts selected for habit-formation study: Group PAL (P1-P5) using visual contexts, and Group Control (P6-P10) using only non-visual contexts.

	Desired contexts	Chosen intervention context
P1	*"in room with partner"*	Custom Face[partner] AND Indoor Location[room]
P2	*"brushing teeth at night"*	Custom Context[Brushing Teeth] AND Time[8–9]
P3	*"phone/computer"*	Contexts = Object[Phone] OR Object[Computer]
P4	*"leaving lab in evening"*	Time[5–7] AND Indoor Location[lab exit]
P5	*"train station in morning"*	Context Cluster[train station] AND Time[9–11]
P6	*"dinner"*	Time[9–10] AND Location[home address]
P7	*"entering dorm room"*	Location[Dorm address]
P8	*"leaving home in morning"*	Time[8–10] AND Location = [home address]
P9	*"in evening at work"*	Time[5pm–7pm] AND Location = [office]
P10	*"in the morning"*	Time[10am–11am]

Weekly Experience Questionnaire. Finally, at the end of the 4-week study, we conducted an open-ended interview with the participants.

6 Results

Our 4-week study, plus post-study evaluations 1 and 10 weeks later, compared the efficacy of habit-support interventions using visual contexts with those using only non-visual contexts. We summarize the intervention contexts, quantitative habit formation, and qualitative experiences of our participants below.

6.1 Chosen Habit-Support Contexts

3 out of 5 participants (P6-8) in the control group could have used interventions in visual contexts. However, limited by only non-visual contexts, the participants chose an approximation, e.g., *Time[9–10] AND location[home]* for *"dinner"*. All Group PAL participants selected visual contexts and 4 out of 5 trained custom visual contexts, i.e., activities, faces, or indoor locations. The intervention contexts chosen and desired by each participant are in Table 3.

6.2 Quantitative Habit Formation

The 4-week increase in SRBAI was 77.1% (Group PAL) and 39.3% (Group Control), and the increase in SRHI was 75.7% (Group PAL) and 21.9% (Group Control). The average week-by-week change for the 4 weeks during the study was: {SRBAI = {Group PAL: [64%, 17%, −13%, 5.5%]; Control: [33%, 14%, −6%, −2.1%]}, and SRHI = {Group PAL: [59%, 21%, −7.9%, −1%]; Control: [26%, 2.3%, −7.3%, and 2.0%]}. Most of the changes in SRHI/SRBAI occurred in the first 2 weeks. Week 3 even had a decrease in SRHI/SRBAI since classes started in Week 3 and our participants, all of whom were students, mentioned

getting 'busy'. SRHI/SRBAI remained relatively stable or in Week 4 and also in the post-study surveys 1 and 10 weeks later. The results are shown in Fig. 4.

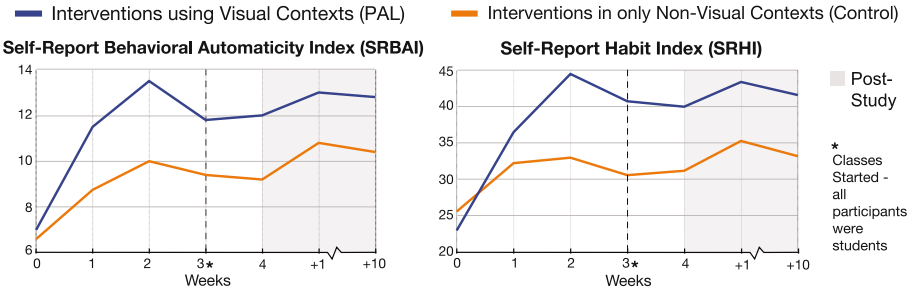

Fig. 4. Habit-formation results for study, N = 10, comparing interventions using visual contexts (Group PAL) with interventions in only non-visual contexts (Group Control).

6.3 Qualitative Responses

We noted the following 3 themes in our questionnaire and interview responses.

Intervention Contexts. Group PAL found interventions in the right contexts helpful (*P1: "reminders at the right time were helpful", P3: "notifications while in front of my laptop to not mindlessly drift into work"*), especially when they were busy (*P4: "reminders especially as I got busy", P2: "I did my habit even though I was busy!"*). The control group did not find the interventions in only non-visual contexts as helpful because they were not in the right moments (*P8: "reminders were not for the exact moments I wanted", P6: "notifications at general times can be anxiety-inducing."*) – some ignored the notifications (*P9: "I did not notice the reminders because I was busy"*), while others used them as persistent, not just-in-time, reminders (*P10: "notification kept reminding me about focus(ing) on a healthy life"*).

System Usability. The participants did not have privacy or social acceptability problems with the camera. Some participants mentioned liking the camera (*P2: "liked the small and unnoticeable camera", P3: "nice to be able to cover the camera with hair when needed"*), while others had minor complaints (*P5: "headphones and winter cap cover the camera"*) or suggestions (*P4: "a hardware switch to turn off the camera in restrooms"*). Moreover, three participants found the device 'heavy' and one of them even mentioned that (*"P4: I didn't wear the device for very long due to its bulkiness"*), whereas two participants complained about the limited battery life (*"P1: battery doesn't last full day"*). Lastly, four participants had minor issues due to mobile app crashes and battery drain.

Emotions and Self-perceptions. After the study, Group PAL participants, in general, indicated a firmer belief in their ability to change (*"P2: (I learned) I can take out time for activities I thought there wasn't time for", P5: "being able to do what I had planned to do gave me the confidence to change", P3: "As I practiced more, it became a part of my day"*), compared to the control group participants, two of whom were apprehensive even after they were able to change (*P5: "I was successful, I am happy but I am also worried that I'll be able to keep it going."*). Group PAL participants felt good (*P1: "I was successful...felt really good", P4: "healthy lifestyle doesn't only keep you physically fit but also happy and confident"*), but the control group did not mention anything explicitly.

7 Discussion

We developed a wearable system for just-in-time habit-support interventions in personalized visual contexts, and used it to compare the efficacy of habit-support interventions using visual contexts with those using only non-visual contexts. We leveraged deep learning to extend real-world habit-support interventions to wearable egocentric visual contexts and our study shows that interventions using personalized egocentric visual contexts can support better real-world habit formation for behavior change than the existing habit-support interventions in only non-visual contexts, i.e., time, geolocation, and physical activity. We discuss the key findings, limitations, and recommendations of our study below.

7.1 Key Findings

We summarize our key findings for quantitative and qualitative results for interventions using visual contexts versus interventions in only non-visual contexts.

Quantitative Habit Formation. Our 4-week study with total 10 participants showed almost double habit formation with interventions using personalized wearable egocentric visual contexts than with interventions in only non-visual contexts. The habits also persisted in the post-study surveys 1 and 10 weeks later, showing sustainable habits without long-term dependence on app support.

Interventions in Visual Contexts. All Group PAL participants selected interventions in at least one visual context, and though we did not measure the context detection accuracy in our 4-week study due to privacy reasons, the participants mentioned receiving interventions in helpful contexts. Even though cameras usually have privacy concerns, Group PAL's participants did not mention any because their images were not sent to the cloud/another device for processing and all the images for custom labeling were also deleted right after the users labeled them. Overall, the participants found interventions using visual contexts timely and useful, even during the participants' busy days, and the camera was usable because of its small size and proper data privacy and control measures.

Interventions in only Non-visual Contexts. Some participants wanted contexts, e.g., "dinner", "leaving home", and "entering dorm room", which were not only perfectly recognizable using non-visual contexts, i.e., time, geolocation, and physical activity, and could have been better recognized by adding visual context detection. Thus, the participants did not receive interventions in their exact desired contexts an had to either remember to do their habits in their exact desired contexts or do the habits based on the non-visual context interventions.

Moreover, the participants were not always doing the same activity in only time or geolocation-based contexts, and doing the desired habits when they received the intervention meant that their actions were not in stable contexts, e.g., the participant could be having dinner or working at 9 pm. Habits form in stable contexts and unstable contexts may have hindered habit formation.

Finally, interventions in only non-visual contexts were disruptive and even anxiety-inducing because the exact activities the participants were doing in the non-visual contexts varied and it was not ideal to disrupt them. Thus, the participants ignored the interventions and even skipped habits when they were busy.

7.2 Limitations and Future Work

Our work has the following three limitations and directions for future work.

Device Usability and Functionality. Since our device was a lab-made prototype, it was relatively bulky and had to be recharged for day-long use. However, with industrial design and manufacturing as well as further advancements in on-device deep learning hardware and models, the wearable device could be made much smaller and also have a longer battery life. Also, further research into on-device deep learning models could open up new possibilities for visual context detection, including human-in-the-loop personalized visual context detection.

Study Size and Participants. Our study is an initial investigation into using visual context detection for habit-support interventions. We kept our study group small to evaluate the detailed experience of each user. Future iterations of our work may involve larger-scale studies with more participants, and also, potentially more diverse groups, e.g., people with specific behavior change needs.

Behavior Change Measurement. We did not measure actual behavior change using sensor data or self-report because of privacy reasons and because we did not want to put implicit pressure on the participants to change their behaviors, knowing that they were being monitored or had to self-report. Instead, we used habit formation as an efficacy measure since our system was designed for habit support [15]. In the future, visual context sensing can be extended to objectively track behavior change and even offer closed-loop behavior change interventions.

7.3 Recommendations and Implications

We have three main recommendations. First, consider including personalized visual contexts as users want interventions in personalized visual contexts. Second, use on-device deep learning to keep user data private. Third, use visual context detection to deliver habit-support interventions for better habit formation. PAL supports better real-world habit formation using egocentric visual contexts, and can also be further useful for privacy-preserving visual context tracking and for non-habit-support behavior change interventions in egocentric visual contexts.

8 Conclusion

Habit formation helps sustain behavior change [17,40]. We investigated if adding egocentric visual contexts to the existing non-visual mobile contexts can improve context-aware habit-support interventions in the real world. We conducted a user survey about the desired habit-support intervention contexts and mediums. Based on our survey, we created a wearable system, named PAL, to deliver habit-support interventions in personalized visual contexts, while preserving user privacy using on-device deep learning. Our study shows that using personalized visual contexts for context-aware habit-support interventions leads to more habit formation than interventions using only non-visual mobile contexts. The habits also persisted 10 weeks after the study. Thus, PAL's wearable interventions in egocentric visual contexts improve real-world context-aware habit-support interventions for better habit formation and sustainable behavior change.

Appendix: Model Evaluation Data

We share below additional details about the in-the-wild data collected for evaluating our machine learning models.

1. Overall: 13 locations (9 indoors - 4 eateries, 2 shops, 1 dorm, 1 house, 1 office; 4 outdoors - 1 shopping area, 1 roadside walkway, 1 train station, 1 residential area).
2. Object detection: 618 persons, 282 books, 48 TV screens, 45 laptops, 30 chairs, 25 bottles, 14 cars, 13 teddy bears, 8 keyboards, 7 microwaves, 7 cell phones, 6 potted plants, 5 couches, 4 bowls, 3 sandwiches, 3 trains, 2 clocks, 2 refrigerators, 2 sinks, 2 dining tables, 1 toilet, 1 umbrella, 1 bus, and 1 bicycle.
3. Custom activities: brushing teeth, making coffee, eating lunch, working in own office, working in an open office area, playing pool, playing foosball (∼50 images each) (Fig. 5).

Fig. 5. Example images, N = ∼1000, from in-the-wild evaluations of on-device models.

References

1. Bauer, Z., Dominguez, A., Cruz, E., Gomez-Donoso, F., Orts-Escolano, S., Cazorla, M.: Enhancing perception for the visually impaired with deep learning techniques and low-cost wearable sensors. Pattern Recogn. Lett. **137**, 27–36 (2019)
2. Costa, J., Adams, A.T., Jung, M.F., Guimbretière, F., Choudhury, T.: EmotionCheck: a wearable device to regulate anxiety through false heart rate feedback. GetMobile **21**(2), 22–25 (2017)
3. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
4. Dimiccoli, M., Marín, J., Thomaz, E.: Mitigating bystander privacy concerns in egocentric activity recognition with deep learning and intentional image degradation. Proc. ACM Interact. Mob. Wearable Ubiquit. Technol. **1**(4), 1–18 (2018)
5. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD, vol. 96, pp. 226–231 (1996)
6. Fogg, B.: A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology, Persuasive 2009, pp. 40:1–40:7. ACM, New York (2009). https://doi.org/10.1145/1541948.1541999
7. Gardner, B., Abraham, C., Lally, P., de Bruijn, G.J.: Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the self-report habit index. Int. J. Behav. Nutr. Phys. Act. **9**(1), 102 (2012)

8. Gollwitzer, P.M.: Implementation intentions: strong effects of simple plans. Am. Psychol. **54**(7), 493 (1999)
9. Hardeman, W., Houghton, J., Lane, K., Jones, A., Naughton, F.: A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int. J. Behav. Nutr. Phys. Act. **16**(1), 31 (2019)
10. Hekler, E.B., Klasnja, P., Froehlich, J.E., Buman, M.P.: Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2013, pp. 3307–3316. ACM, New York (2013)
11. Howard, A.G., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications, April 2017
12. IJsselsteijn, W., de Kort, Y., Midden, C., Eggen, B., van den Hoven, E.: Persuasive technology for human well-being: setting the scene. In: IJsselsteijn, W.A., de Kort, Y.A.W., Midden, C., Eggen, B., van den Hoven, E. (eds.) PERSUASIVE 2006. LNCS, vol. 3962, pp. 1–5. Springer, Heidelberg (2006). https://doi.org/10.1007/11755494_1
13. Judah, G., Gardner, B., Aunger, R.: Forming a flossing habit: an exploratory study of the psychological determinants of habit formation. Br. J. Health. Psychol. **18**(2), 338–353 (2013)
14. Kaur, H., Williams, A.C., McDuff, D., Czerwinski, M., Teevan, J., Iqbal, S.T.: Optimizing for happiness and productivity: modeling opportune moments for transitions and breaks at work. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–15 (2020)
15. Klasnja, P., Consolvo, S., Pratt, W.: How to evaluate technologies for health behavior change in HCI research. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2011, pp. 3063–3072. ACM, New York (2011)
16. Kuznetsova, A., et al.: The open images dataset v4: unified image classification, object detection, and visual relationship detection at scale. arXiv preprint arXiv:1811.00982 (2018)
17. Kwasnicka, D., Dombrowski, S.U., White, M., Sniehotta, F.: Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol. Rev. **10**(3), 277–296 (2016)
18. Lane, N.D., Bhattacharya, S., Mathur, A., Georgiev, P., Forlivesi, C., Kawsar, F.: Squeezing deep learning into mobile and embedded devices. IEEE Pervasive Comput. **16**(3), 82–88 (2017)
19. Lee, E., Lee, W., Cho, J.: Moonglow: wearable device which helps with cognitive behavioral therapy for panic disorder patients. In: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, UbiComp 2018, pp. 126–129. ACM, New York (2018)
20. Lee, H., Upright, C., Eliuk, S., Kobsa, A.: Personalized object recognition for augmenting human memory. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, pp. 1054–1061 (2016)
21. Lee, H., Upright, C., Eliuk, S., Kobsa, A.: Personalized visual recognition via wearables: a first step toward personal perception enhancement. In: Costa, A., Julian, V., Novais, P. (eds.) Personal Assistants: Emerging Computational Technologies. ISRL, vol. 132, pp. 95–112. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-62530-0_6
22. Lee, J., Walker, E., Burleson, W., Kay, M., Buman, M., Hekler, E.B.: Self-experimentation for behavior change: design and formative evaluation of two

approaches. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 6837–6849 (2017)
23. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
24. Mathur, A., Lane, N.D., Bhattacharya, S., Boran, A., Forlivesi, C., Kawsar, F.: Deepeye: resource efficient local execution of multiple deep vision models using wearable commodity hardware. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, pp. 68–81 (2017)
25. McEuen, A., Proffitt, J., Camba, J.D., Kwon, E.S.: &You: design of a sensor-based wearable device for use in cognitive behavioral therapy. In: Duffy, V.G., Lightner, N. (eds.) Advances in Human Factors and Ergonomics in Healthcare. AISC, vol. 482, pp. 251–260. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-41652-6_24
26. Nahum-Shani, I., et al.: Just-in-Time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann. Behav. Med. **52**(6), 446–462 (2018)
27. Nishajith, A., Nivedha, J., Nair, S.S., Shaffi, J.M.: Smart cap-wearable visual guidance system for blind. In: 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 275–278. IEEE (2018)
28. Pinder, C., Vermeulen, J., Cowan, B.R., Beale, R.: Digital behaviour change interventions to break and form habits. ACM Trans. Comput. Hum. Interact. **25**(3), 15:1–15:66 (2018)
29. Pinder, C., Vermeulen, J., Wicaksono, A., Beale, R., Hendley, R.J.: If this, then habit: exploring context-aware implementation intentions on smartphones. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, pp. 690–697. ACM, September 2016
30. Qi, H., Brown, M., Lowe, D.G.: Low-shot learning with imprinted weights. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5822–5830 (2018)
31. Renfree, I., Harrison, D., Marshall, P., Stawarz, K., Cox, A.: Don't kick the habit: the role of dependency in habit formation apps. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2932–2939 (2016)
32. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)
33. Stawarz, K., Cox, A.L., Blandford, A.: Don't forget your pill! designing effective medication reminder apps that support users' daily routines. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2269–2278 (2014)
34. Stawarz, K., Cox, A.L., Blandford, A.: Beyond self-tracking and reminders: designing smartphone apps that support habit formation. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015, pp. 2653–2662. ACM, New York (2015)
35. Tikka, P., Oinas-Kukkonen, H.: RightOnTime: the role of timing and unobtrusiveness in behavior change support systems. In: Meschtscherjakov, A., De Ruyter, B., Fuchsberger, V., Murer, M., Tscheligi, M. (eds.) PERSUASIVE 2016. LNCS, vol. 9638, pp. 327–338. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-31510-2_28

36. Verplanken, B., Orbell, S.: Reflections on past behavior: a self-report index of habit strength 1. J. Appl. Soc. Psychol. **33**(6), 1313–1330 (2003)
37. Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E.: Deep learning for computer vision: a brief review. Comput. Intell. Neurosci. **2018** (2018)
38. Wicaksono, A., Hendley, R., Beale, R.: Investigating the impact of adding plan reminders on implementation intentions to support behaviour change. Interact. Comput. **31**(2), 177–191 (2019)
39. Wicaksono, A., Hendley, R.J., Beale, R.: Using reinforced implementation intentions to support habit formation. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–6 (2019)
40. Wood, W., Neal, D.T.: Healthy through habit: interventions for initiating & maintaining health behavior change. Behav. Sci. Policy **2**(1), 71–83 (2016)