Towards AI-Powered AR for Enhancing Sports Playability for People with Low Vision: An Exploration of ARSports

Jaewook Lee[1] Yang Li[1] Dylan Bunarto[1] Eujean Lee[1] Olivia H. Wang[1] Adrian Rodriguez[1]
Yuhang Zhao[2] Yapeng Tian[3] Jon E. Froehlich[1]

University of Washington[1] University of Wisconsin-Madison[2] University of Texas at Dallas[3]

Figure 1: We contribute *ARSports*, a near real-time wearable stereo AR system to enhance sports playability for people with low vision composed of: (1) first-person basketball and tennis image datasets, which we manually collected and annotated; (2) accompanying fine-tuned instance segmentation models; and (3) a wearable AR research prototype that overlays visual augmentations (*i.e.,* instance segmentation masks) in an LV user's residual field-of-view.

Abstract

People with low vision (LV) experience challenges in visually tracking balls and players in sports like basketball and tennis, which can adversely impact their participation and health. We introduce *ARSports*, a wearable AR research prototype that overlays instance segmentation masks in near real-time for improving sports accessibility. To create ARSports, we manually collected and annotated novel first-person perspective sports datasets, fine-tuned instance segmentation models using this labeled data, and built an initial wearable AR prototype by combining the *ZED Mini* stereo camera with the *Oculus Quest 2* VR headset. Our evaluations suggest that combining real-time computer vision and augmented reality to create scene-aware visual augmentations is a promising approach to enhancing sports participation for LV individuals. We contribute open-sourced egocentric basketball and tennis datasets and models, as well as insights and design recommendations from our pilot study with an LV research team member.

Index Terms: augmented reality, accessibility, visual augmentation, computer vision, sports

1 Introduction

Low vision (LV) individuals face unique challenges in sports and exercise, which can negatively impact participation as well as physical and mental health [1, 11, 13, 22, 3, 32, 30]. Competitive ball-based sports like basketball, tennis, and soccer, for example, involve fast-moving objects such as balls and players that are difficult to visually identify and track [30]. However, prior HCI studies have largely focused on improving how blind or low vision (BLV) people watch sports [26, 8, 14, 15, 27], rather than enabling their participation. While some research has examined camera-based assistance for casual non-ball games [18, 19] and exercise video games introducing ball sports to LV people [25, 24, 29, 28, 31], a significant gap remains in developing real-time wearable AR-assisted vision to empower LV individuals to play ball sports independently [20].

In this paper, we explore how augmented reality (AR) glasses and computer vision (CV) may enable broader sports participation for LV individuals. We introduce *ARSports*, a wearable AR prototype for supporting LV people to play ball-based sports using near real-time instance segmentation and visual augmentation (Figure 1). To create ARSports, we: (1) manually collected and labeled novel egocentric basketball and tennis image datasets; (2) fine-tuned instance segmentation models on these datasets; and (3) constructed a wearable stereo AR prototype capable of displaying visual augmentations in near real-time (∼20-25 FPS). Our work bridges the fields of HCI and CV: existing sports datasets [6, 7, 21] and models [12, 10, 33] are often designed for third-person views, and current off-the-shelf AR headsets like the *Microsoft HoloLens 2* do not support long-range real-time depth sensing, both of which have prevented the use of AI-powered AR in sports scenarios.

To address these limitations, we first recorded first-person point-of-view basketball and tennis 1080p@30fps videos using a HoloLens 2 headset. We then selected critical frames with YOLOv8 [16] followed by manual filtering, and labeled them using *Roboflow*[1] equipped with the *Segment Anything Model* (SAM) [17]. This resulted in a dataset of 5,412 egocentric sports images (2,430 basketball images and 2,982 tennis images). We then fine-tuned *RT-MDet* [5], a state-of-the-art instance segmentation model[2] on our

[1]https://roboflow.com
[2]http://tinyurl.com/sota-instance-segmentation

datasets. Finally, to help LV sports players identify key objects, we implemented near real-time AR visual augmentations using an *Oculus Quest 2* virtual reality (VR) headset[3] combined with a *ZED Mini* stereo camera[4] for depth estimation.

To evaluate our approach, we conducted two studies: first, a technical performance evaluation of our fine-tuned segmentation models, which shows the effectiveness of fine-tuning for targeted tasks versus the base RTMDet model. Then, we conducted a pilot evaluation with an LV research team member on actual basketball and tennis courts. He interacted with our wearable AR prototype for 30 minutes per sport, then shared his feedback and design suggestions. Preliminary findings suggest that ARSports is effective in helping LV people visually perceive various sports elements such as players, balls, and nets. Our LV research team member emphasized the need for simple designs to ensure visual augmentations do not interfere with an LV person's remaining visual field-of-view.

In summary, our key contributions include: (1) open-sourced first-person basketball and tennis image datasets, as well as accompanying fine-tuned instance segmentation models; (2) a research prototype for wearable AR capable of tracking and visually augmenting different elements of basketball and tennis such as balls and players; and (3) findings from a pilot evaluation with an LV research team member. To enable others to build off our work, we open-sourced our datasets and models here: https://github.com/makeabilitylab/ARSports.

2 System Implementation

We first explain our methods for collecting and annotating egocentric image datasets in basketball and tennis, followed by how we fine-tune instance segmentation models to these datasets and generate visual augmentations in stereo AR. By augmenting LV people's residual field-of-view with instance segmentation results, we aim to enhance the visual saliency of different sports elements to provide a better sense of shape, contour, location, and depth.

2.1 Data Collection and Annotation

To address the lack of first-person sports recordings, we first manually assembled egocentric sports datasets by collecting and annotating video recordings captured from a first-person perspective.

2.1.1 Data Collection

Our custom ARSports image datasets currently feature two sports: basketball and tennis, which were selected due to their popularity and the presence of fast-moving elements such as balls and players.

We instrumented a player with a *Microsoft HoloLens 2* headset to actively engage in each sport and collect video recordings (1080p@30fps). We chose a commercially-available AR headset over more sophisticated cameras like a *GoPro* because the latter produces high-resolution, stabilized video that does not accurately represent the video capture capabilities of AR headsets and user's constant head motion when playing sports. For basketball, a player wearing the HoloLens performed various common tasks such as *shooting*, *passing*, *dribbling*, *defending*, and *being defended* in an indoor 3 *vs.* 3 basketball game. Tennis data was collected from three 1 *vs.* 1 rallies, where players executed *ground strokes*, *volleys*, *serves*, and *return of serves*. We then carefully trimmed irrelevant parts from the recordings, such as the player interacting with the HoloLens to start and stop video capture, and compiled the clips into approximately an hour of footage for each sport.

To extract images from the finalized basketball and tennis video footage, we first utilized YOLOv8 [16] to find frames with a sports ball, skipping 20 frames each time one is found to reduce redundancy. Then, we manually removed repetitive, excessively blurry, and non-informative frames. This process resulted in a total of 1,431 first-person basketball images and 1,754 first-person tennis images. Lastly, we blurred people's faces using the *CenterFace* algorithm [34] to ensure anonymity.

2.1.2 Data Annotation

We labeled the extracted frames using *Roboflow*[1], an online tool for annotating, training, and optimizing CV models, and the included *Segment Anything Model* (SAM) [17]. For basketball, we labeled: *people*, *basketball*, *hoop*, and *backboard*. For tennis, we labeled: *people*, *tennis ball*, *net*, and *racket*. The polygon annotations were initially done by SAM, which we then adjusted manually. We empirically ignored objects that were too blurry to label.

To handle occlusion (*e.g.,* a basketball can obscure parts of a person or a tennis net can cover a person's leg), we employed the following heuristic (Figure 3): (1) an object fully split by another object is treated as a single annotation with correct layering (*e.g.,* if a basketball fully splits a person, the person is labeled using one polygon annotation, and the basketball is in a layer above the person); (2) parts of an object that are occluded by another object but still visible are included in the annotation (*e.g.,* a tennis player's legs are often behind the net, but still visible, and so the legs are included in the person's polygon annotation); and (3) parts at the ends of an object fully occluded by another object are not included in the annotation (*e.g.,* if a tennis player's shoes are not visible because they line up with the top of the net, then the person's polygon annotation stops at their ankles).

After annotating, discussing, and cross-checking amongst the research team, we applied several image transformations including crop with 0% minimum zoom and 40% maximum zoom, rotation between -15° and +15°, brightness between -15% and +15%, blur up to 2.5px, and noise up to 0.1% of pixels, and the images were adjusted to fit a 640x480 resolution (*i.e.,* MS COCO [23] average image resolution), resulting in the final dataset of 5,412 total images: 2,430 basketball and 2,982 tennis.

2.2 Fine-tuning an Instance Segmentation Model

As we aim to display visual augmentations in real-time, we experimented with approaches that can deliver both speed and accuracy. We chose to fine-tune the *RTMDet* model [5], specifically its *RTMDet-Ins-l* variant, on our datasets, as it is a state-of-the-art real-time instance segmentation model[2] trained on the MS COCO dataset [23], promising accuracy up to 43.7% mask AP and speeds up to 271 FPS on an NVIDIA 3090 GPU.

2.2.1 Fine-tuning RTMDet

We chose to fine-tune RTMDet rather than train it from scratch to make it work for our smaller, class-specific dataset. To achieve this, we utilized the fine-tuning pipeline provided by the *MMDetection* library [4], a PyTorch-based open-source toolbox for object detection. We began with a pre-trained RTMDet-Ins-l model, froze its backbone, modified the model configuration file to match our label classes, and then trained it on our basketball and tennis datasets. These customized models, dubbed *RTMDet-Ins-l-Basketball* and *RTMDet-Ins-l-Tennis* respectively, underwent training for 150 epochs using a batch size of 4 on a single CUDA-enabled NVIDIA 4080 GPU. We show inferencing results of RTMDet-Ins-l-Basketball and RTMDet-Ins-l-Tennis on images from the validation subset of our datasets in Figure 2. We also open-sourced our **datasets and model weights**, as well as our **fine-tuning steps**, giving researchers the tools to expand ARSports.

2.2.2 Model Evaluation

To evaluate RTMDet-Ins-l-Basketball and RTMDet-Ins-l-Tennis, we compared their performance against the base RTMDet-Ins-l on our sports datasets. We used MMDetection's [4] model testing

[3]https://www.meta.com/quest/products/quest-2/
[4]https://store.stereolabs.com/products/zed-mini

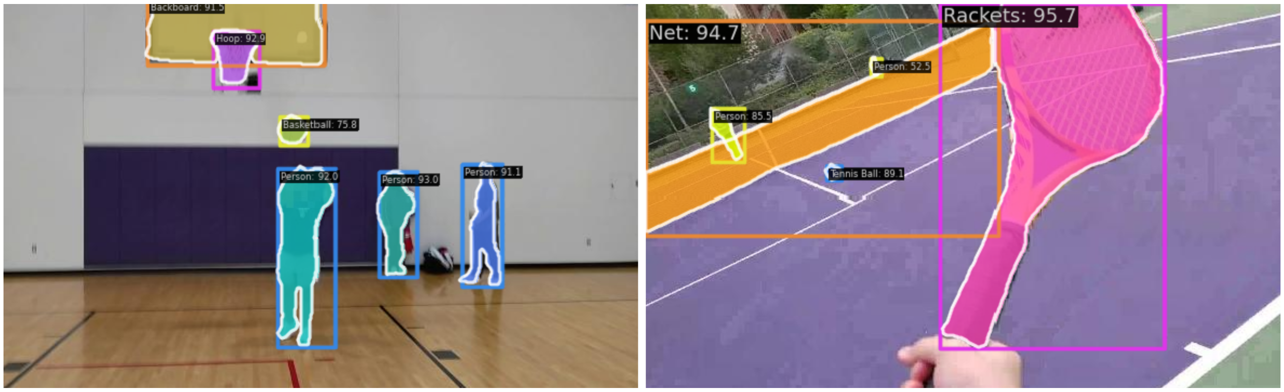

Figure 2: Example inference results of RTMDet-Ins-l-Basketball (left) and RTMDet-Ins-l-Tennis (right) on validation images from our datasets.

(1) Fully Split **(2) Occluded but Visible** **(3) Fully Occluded**

Figure 3: Our annotation heuristic: (1) A basketball fully splitting a person is labeled with the person as one polygon and the basketball on a layer above, (2) A net occluding parts of a visible player is labeled with the person as one polygon and the net on a layer above, (3) A person's feet are fully occluded, so the person polygon stops at the ankles.

Model Name	mAP	AP@50	AP@75
RTMDet-Ins-l (COCO)	0.437	0.660	0.470
RTMDet-Ins-l (tennis dataset)	0.284	0.483	0.266
RTMDet-Ins-l-Tennis	**0.419**	**0.656**	**0.37**
RTMDet-Ins-l (basketball dataset)	0.211	0.348	0.219
RTMDet-Ins-l-Basketball	**0.569**	**0.878**	**0.576**

Table 1: Evaluation of our fine-tuned instance segmentation models. *RTMDet-Ins-l-Tennis* and *RTMDet-Ins-l-Basketball* achieve superior performance across all metrics on our egocentric sports datasets, outperforming the state-of-the-art *RTMDet-Ins-l* model. For reference, we also include *RTMDet-Ins-l* results on the COCO dataset.

pipeline, which conducts evaluations using the test subset of a given dataset. With Roboflow, we generated a test set of 145 basketball and 175 tennis images with an 82-12-6 train-validation-test split.

Accuracy in instance segmentation tasks is typically assessed 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) [9]. *IoU*, integral to these metrics, measures the overlap between predicted segmentation masks and the ground truth, providing a direct indication of spatial alignment accuracy. Both RTMDet-Ins-l-Basketball and RTMDet-Ins-l-Tennis outperform the RTMDet-Ins-l baseline on our sports datasets across mAP, AP@50, and AP@75 (See Table 1).

2.3 Generating Visual Augmentations

With our fine-tuned models, we built a wearable stereo AR prototype that can overlay instance segmentation masks on top of sports elements in near real-time (∼20-25 FPS). To generate visual augmentations in 3D space, we built a custom stereo video-see-through AR system by combining a *ZED Mini* stereo camera[4] with an *Oculus Quest 2* VR headset[3], as current AR headsets like the Microsoft HoloLens 2 do not support long-range real-time depth sensing. Our research prototype streams image frames to an external server over TCP, performs instance segmentation using our fine-tuned models, converts the resulting JSON into a *Protocol Buffers* message[5], streams this message back to ZED, deserializes the message back into JSON, creates ZED-compatible textures (colored overlays), and performs depth estimation to position the visual augmentations. See Figure 4.

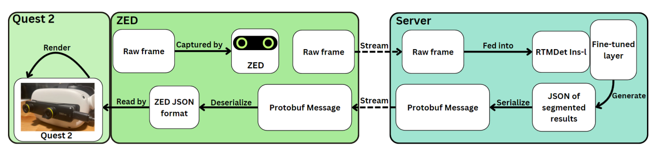

Figure 4: System diagram of ARSports showing how data flows between an AR headset and an external server.

3 Pilot Evaluation

To further evaluate our models and research prototype, we conducted a pilot study with an LV research team member, who played basketball and tennis while wearing our system. He has no light perception in his left eye and a visual acuity of 20/400 in his right eye. He played each sport for 30 minutes (Figure 5), then provided feedback regarding the usability and design considerations of a wearable AR system aimed at enhancing sports playability for LV people. We report preliminary findings below.

Overall, our LV research team member highlighted ARSports as "*effective*," "*helpful*," "*reasonably fluid*," and "*full of potential*." Despite technical challenges such as latency and inconsistent tracking, ARSports is the most advanced AR and CV solution he has tried for fast-paced tasks like playing basketball and tennis. He envisions that a system like ARSports will promote sports participation among LV individuals by empowering them to better visually perceive balls, teammates, and relevant sports equipment. From a

[5]https://protobuf.dev

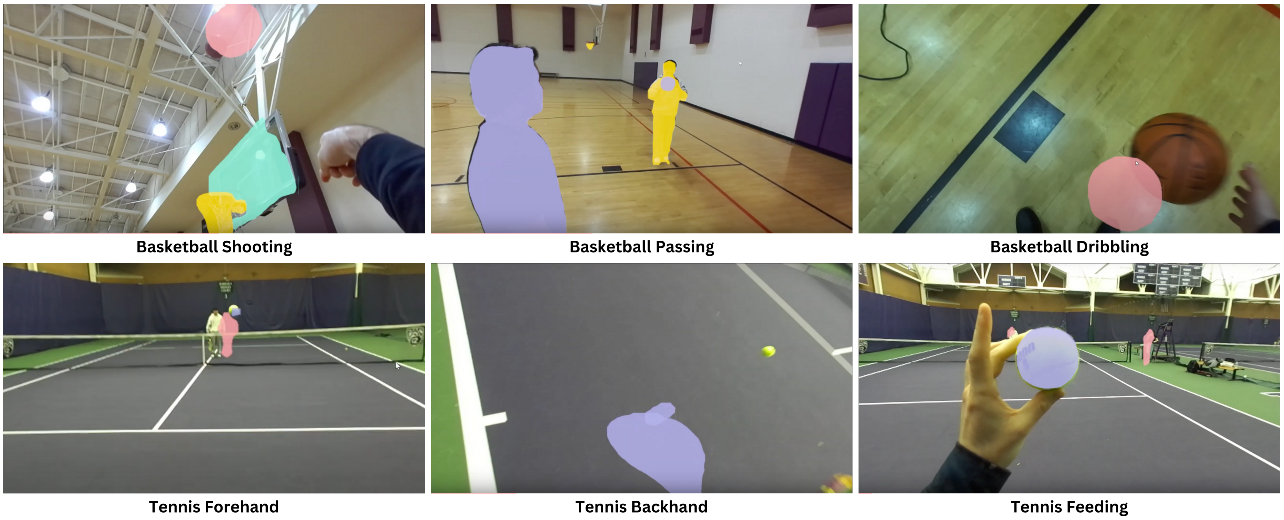

Figure 5: Example images from the first-person view of our LV research team member playing basketball and tennis while using ARSports. We cover different plays that commonly occur in each sport.

Figure 6: Proposed designs sketched by our LV research team member. He emphasized simple designs with customization options.

design perspective, displaying whole instance segmentation masks for sports components is impractical, as they obstruct crucial parts of objects and his remaining field-of-view. For example, applying a solid polygon mask over the basketball hoop and net impedes the view into the hoop. In tennis, a large net polygon obscures most of his visual field, hence we had to disable it during our evaluation.

Our LV research team member suggested two primary improvements: (1) creating simple visual augmentations to lower occlusion and cognitive load; and (2) maximizing user customization. When playing tennis with ARSports, he noted "*Desaturating large graphics preserves general visibility for me. For example, the tennis net perhaps shouldn't be highlighted entirely because it covers too much of my remaining field-of-view. Instead, a line at the top of the net is sufficient for understanding how high I need to hit the ball to make it over the net.*"

To control how much screen space visual augmentations should cover, he suggested defining "*visual pressure*" of rendered graphics: "*I recommend defining a measure for 'visual pressure', which could be the 'total weight' of rendered graphics on screen or even total augmented pixels for a given frame. This serves as a minimum and maximum for rendering amount.*" Additionally, he emphasized the importance of customization options like colors, which shapes to render, and visual pressure threshold. "*For example, in tennis, users with an acuity of 20/200 may benefit from seeing the silhouettes of other players. However, those with an acuity of 20/800 may benefit from an even more abstracted depiction of others, such as rectangular estimates of key features like head and racket. Not everyone needs perfect polygon segmentation masks.*"

He concluded by saying "*I think simplicity actually affords the most utility for low vision people.*" He then sketched design recommendations, which we converted to mockups in Figure 6.

4 Future Work and Conclusion

In this paper, we introduce ARSports, a significant advancement over prior work in aiding low vision sports play via real-time CV and visual augmentations. We contribute first-person perspective basketball and tennis image datasets, instance segmentation models fine-tuned on these datasets, and a wearable AR research prototype that overlays visual augmentations in an LV person's residual field-of-view. A preliminary evaluation with an LV research team member suggests that merging CV and AR technologies can effectively enhance the playability of sports for LV individuals, but should be carefully designed to not add visual clutter.

For future work, (1) our RTMDet-Ins-l-Tennis model occasionally fails to detect tennis balls, highlighting the need for a larger dataset [2] and the selection of more suitable models; (2) we need improved real-time object tracking and depth sensing to ensure more consistent visual augmentations; and (3) we need to study a wider range of augmentation designs, from basic shapes and outlines to more intricate masks, to accommodate users' diverse vision levels. We invite the community to explore ways to improve first-person sports playability for people with different abilities.

Acknowledgments

This work was primarily supported by an NSF GRFP.

References

[1] N. R. Ayvazoglu, H.-K. Oh, and F. M. Kozub. Explaining physical activity in children with visual impairments: A family systems approach. *Exceptional Children*, 72(2):235–248, 2006. 1

[2] M. Bock, H. Kuehne, K. Van Laerhoven, and M. Moeller. Wear: An outdoor sports for wearable and egocentric activity recognition. *CoRR*, abs/2304.05088, 2023. 4

[3] M. Capella-McDonnall. The need for health promotion for adults who are visually impaired. *Journal of Visual Impairment & Blindness*, 101(3):133–145, 2007. 1

[4] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C. C. Loy, and D. Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019. 2

[5] L. Chengqi, Z. Wenwei, H. Haian, Z. Yue, W. Yudong, L. Yanyi, Z. Shilong, and C. Kai. Rtmdet: An empirical study of designing real-time object detectors. *arXiv:2212.07784*, 2022. 1, 2

[6] Y. Cui, C. Zeng, X. Zhao, Y. Yang, G. Wu, and L. Wang. Sportsmot: A large multi-object tracking dataset in multiple sports scenes, 2023. 1

[7] A. Deliège, A. Cioppa, S. Giancola, M. J. Seikavandi, J. V. Dueholm, K. Nasrollahi, B. Ghanem, T. B. Moeslund, and M. V. Droogenbroeck. Soccernet-v2 : A dataset and benchmarks for holistic understanding of broadcast soccer videos. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2021. 1

[8] C. Goncu and D. J. Finnegan. 'did you see that!?' enhancing the experience of sports media broadcast for blind people. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, and K. Inkpen, eds., *Human-Computer Interaction – INTERACT 2021*, pp. 396–417. Springer International Publishing, Cham, 2021. 1

[9] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017. 3

[10] Y. He, Z. Yuan, Y. Wu, L. Cheng, D. Deng, and Y. Wu. Vistec: Video modeling for sports technique recognition and tactical analysis. *AAAI Conference on Artificial Intelligence*, 2024. 1

[11] W. Hopkins, H. Gaeta, A. Thomas, and P. Hill. Physical fitness of blind and sighted children. *European journal of applied physiology and occupational physiology*, 56(1):69–73, 1987. 1

[12] Y.-C. Huang, I.-N. Liao, C.-H. Chen, T.-U. İk, and W.-C. Peng. Tracknet: A deep learning network for tracking high-speed and tiny objects in sports applications, 2019. 1

[13] J. M. Jacobs, R. Hammerman-Rozenberg, Y. Maaravi, A. Cohen, and J. Stessman. The impact of visual impairment on health, function and mortality. *Aging clinical and experimental research*, 17:281–286, 2005. 1

[14] G. Jain, B. Hindi, C. Courtien, C. Wyrick, X. Y. T. Xu, M. C. Malcolm, and B. A. Smith. Towards accessible sports broadcasts for blind and low-vision viewers. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544549.3585610 1

[15] G. Jain, B. Hindi, C. Courtien, X. Y. T. Xu, C. Wyrick, M. Malcolm, and B. A. Smith. Front row: Automatically generating immersive audio representations of tennis broadcasts for blind viewers. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3586183.3606830 1

[16] G. Jocher, A. Chaurasia, and J. Qiu. Ultralytics YOLO, Jan. 2023. 1, 2

[17] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick. Segment anything. *arXiv:2304.02643*, 2023. 1, 2

[18] M. Kobayashi. A basic inspection of wall-climbing support system for the visually challenged. In *Computers Helping People with Special Needs: 12th International Conference, ICCHP 2010, Vienna, Austria, July14-16, 2010, Proceedings, Part II 12*, pp. 332–337. Springer, 2010. 1

[19] M. Kobayashi and T. Suzuki. Accessibility improvement of leisure sports "mölkky" for visually impaired players using ai vision. In *International Conference on Computers Helping People with Special Needs*, pp. 73–78. Springer, 2022. 1

[20] J. Lee, D. P. Sarda, E. Lee, A. Lee, J. Wang, A. Rodriguez, and J. E. Froehlich. 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*, UIST '23 Adjunct. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3586182.3615815 1

[21] Y. Li, L. Chen, R. He, Z. Wang, G. Wu, and L. Wang. Multisports: A multi-person video dataset of spatio-temporally localized sports actions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13536–13545, October 2021. 1

[22] L. J. Lieberman and E. McHugh. Health-related fitness of children who are visually impaired. *Journal of visual impairment & blindness*, 95(5):272–287, 2001. 1

[23] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pp. 740–755. Springer, 2014. 2

[24] T. Morelli, J. Foley, L. Columna, L. Lieberman, and E. Folmer. Vitennis: a vibrotactile/audio exergame for players who are visually impaired. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games*, FDG '10, p. 147–154. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1822348.1822368 1

[25] T. Morelli, J. Foley, and E. Folmer. Vi-bowling: a tactile spatial exergame for individuals with visual impairments. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '10, p. 179–186. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1878803.1878836 1

[26] H. Ohshima, M. Kobayashi, and S. Shimada. Development of blind football play-by-play system for visually impaired spectators: Tangible sports. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411763.3451737 1

[27] H. Ohshima, M. Kobayashi, and S. Shimada. Improvement of user interface of blind football play-by-play system for visually impaired spectators-tangible sports. In *International Conference on Human-Computer Interaction*, pp. 345–351. Springer, 2023. 1

[28] K. Rector. The development of novel eyes-free exercise technologies using participatory design. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, p. 327–330. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2559206.2559960 1

[29] K. Rector, C. L. Bennett, and J. A. Kientz. Eyes-free yoga: an exergame using depth cameras for blind & low vision exercise. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2513383.2513392 1

[30] K. Rector, L. Milne, R. E. Ladner, B. Friedman, and J. A. Kientz. 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*, ASSETS '15, p. 203–214. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2700648.2809846 1

[31] K. Rector, R. Vilardaga, L. Lansky, K. Lu, C. L. Bennett, R. E. Ladner, and J. A. Kientz. Design and real-world evaluation of eyes-free yoga: An exergame for blind and low-vision exercise. *ACM Trans. Access. Comput.*, 9(4), apr 2017. doi: 10.1145/3022729 1

[32] L. Smith, S. E. Jackson, S. Pardhan, G. F. López-Sánchez, L. Hu, C. Cao, D. Vancampfort, A. Koyanagi, B. Stubbs, J. Firth, et al. Visual impairment and objectively measured physical activity and sedentary

behaviour in us adolescents and adults: a cross-sectional study. *BMJ open*, 9(4):e027267, 2019. 1

[33] N.-E. Sun, Y.-C. Lin, S.-P. Chuang, T.-H. Hsu, D.-R. Yu, H.-Y. Chung, and T.-U. İk. Tracknetv2: Efficient shuttlecock tracking network. In *2020 International Conference on Pervasive Artificial Intelligence (ICPAI)*, pp. 86–91, 2020. doi: 10.1109/ICPAI51961.2020.00023 1

[34] Y. Xu, W. Yan, H. Sun, G. Yang, and J. Luo. Centerface: Joint face detection and alignment using face as point, 2019. 2