iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81422

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

A Comprehensive Review of Edge-AI Powered Assistive Systems for the Visually Impaired: Models, Hardware, and Future Directions

Harish Kunder, Aashish CA, Fawaz Shaikh, Shashank HN, Sujay SP
Department of Artificial Intelligence and Machine Learning Alva's Institute of Engineering and Technology Mangalore, Karnataka, India

Abstract: Since more than 2.2 billion people worldwide suffer from visual impairment, assistive technology is an importantfield of study. Intelligent systems that assist blind and visually impaired (VI) people in their daily lives have been made possible by recent developments in computer vision, natural language processing, and edge artificial intelligence. Edge platforms like the NVIDIA Jetson family offer low latency, privacy, and offline dependability in contrast to cloud-based solutions.

This survey critically evaluates more than fifty commercialand academic systems in three areas: (1) deep learning modelsfor object detection, image captioning, and OCR; (2) assistive applications for navigation, object recognition, and scene under-standing; and (3) embedded hardware that allows for real-time inference under resource constraints. We examine usability gaps incommercialproductsandtrade-offsbetweenenergy,accuracy, and latency.

Strict wearable energy budgets, limited reproducibility, multi- lingualaccessibility,anddatasetbiasaresomeofthemainobsta-cles.Additionallycoveredarepromisingavenueslikemultimodal fusion,voice-basedinteraction,federatedlearning,andculturally diverse datasets. For researchers and practitioners creating the upcoming generation of AI-powered accessibility technologies, this work acts as a road map.

IndexTerms: Computervision,EdgeAI,multimodallearning, object detection, image captioning, assistive technology, NVIDIA Jetson, Blind or visually impaired

I. INTRODUCTION

Atleastonebillionoftheworld's2.2billionvisuallyimpairedindividualsareavoidableoruntreated[1].Whilemodernassistivesystemsincorporatesensorsandvision-based techniquesforricherperception,traditionalaidslikewhite canes and guide dogs only offer limited functionality [4],[5]. Text recognition, object detection, and scene understanding haveallgreatlyimproved withdeeplearning [2],[3],[13],[14]. Natural-language descriptions of complex environments arenowpossiblethankstovision-languagemodels[8].The integration of these capabilities into daily life is demon-strated by commercial tools like Microsoft Seeing AI, Google Lookout, and OrCam MyEye [9]–[11]. In terms of usability, latency, and offline performance, comparative studies show trade-offs between cloud-based apps and wearables with edge capabilities [17],[18].

Researchers are increasingly using edge AI platforms like the Raspberry Pi with Coral TPU, the NVIDIA Jetson Nano, TX2, and Orin Nano to get around the drawbacks of relying too much on the cloud. These gadgets allow offline support, better privacy, and real-time inference with lower latency[12]. Benchmarks demonstrate their efficacy for assistive tasks; however, issues persist, including lack of open-source reproducibility, dataset bias that restricts generalization, and multilingual OCR and speech support [15],[16].

Three areas are the focus of this survey, which summa-rizes research from more than fifty academic and commercial works: (1) assistive systems for navigation, scene understand- ing, and everyday help [7],[18]; (2) multimodal and deep learningmodelsthatdriveOCR,detection,andcaptioning[2],[3],[8],[13],[14],[19]; and (3) edge hardware that allows deployment in resource-constrained environments [12]. The studyhighlightstheopportunitiesanddifficultiesindeveloping the next generation of AI-powered assistive technologies for blind and VI users by looking at approaches, trade-offs, and practical prototypes.

II. BACKGROUND

Over the past 20 years, assistive technologies for the blind and VI have steadily improved, moving from simple sensing devices to complex edge-AI-powered systems.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

To identify obstacles and provide haptic or auditory cues, early solutions usedultrasonicorinfraredsensors.Theyfrequentlyfailedin complexenvironmentsandwereunabletoconveysemantic context,despitebeingeffectiveincontrolledsettings[4],[5].
Fivekeyphasesinthisprogressionaredepictedin Fig.1:

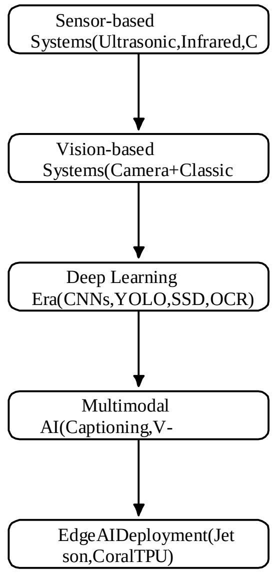

Fig. 1.Timeline of assistive technologies: from sensor-based aids to edge-AIdeployments.

- Sensor-based: Although they lacked object-level recog- nition, infrared/ultrasonic modules allowed for basic ob- stacle alerts.
- Vision-based: Despite introducing depth estimation and lane following, cameras with classical CV struggled in real-world variability [7].
- Deep Learning: Real-time object detection and OCR were made possible by CNNs, YOLO, and SSD, which also supported text access and navigation [2],[3],[13].
- MultimodalAI:Semanticsceneunderstanding(e.g., “bus approaching”) was made possible by vision–language (V-L) models and captioning systems [8].
- Edge AI: By moving inference to the device, gadgetslike the Jetson Nano/TX2/Orin and Coral TPU improved privacy and decreased latency [12].

Reproducibility, multilingual OCR, dataset bias, and hard- ware trade-offs were among the new issues that each stage broughtwhileaddressingpreviousones[15],[16].Thepresent research frontier is now defined by these problems.

ASSISTIVETECHNOLOGIESFORTHEVISUALLYIMPAIRED

Simplesensingmodulesandsophisticated AI-drivenframe- works are examples of assistive technologies. Obstacle alerts were provided by early tools like depth-based devices and ultrasonic canes, but they were unable to record semantic context[4],[5].Ataxonomyofexistingsolutionsisdepicted in Fig. 2, which demonstrates the transition from task-specific designs (OCR, navigation) to multimodal frameworks and commercial products.

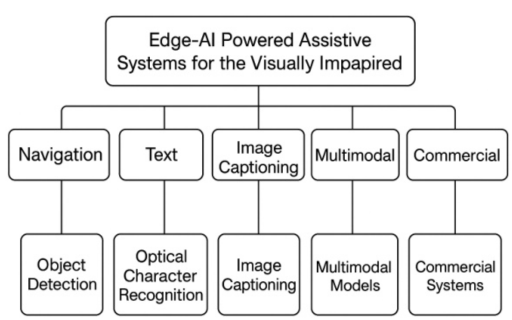

Fig. 2.Taxonomy of Edge-AI powered assistive systems, grouped into navi-gation, text recognition, captioning, multimodal frameworks, and commercialproducts.

Thesystemsfallintothefollowinggeneralcategories:

- Navigation: Mobility and obstacle avoidance are made possible by object detection and depth estimation (e.g., YOLO, SSD) [2],[3],[6]. Their usefulness is confirmed by real-world prototypes like NavWear [18].
- Text Recognition: Currency, product, and document reading are supported by OCR systems (CNN–RNNwith attention). Accessibility is increased by multilingual extensions [13],[14].
- Image Captioning: Captioners improve contextual awareness by using natural language to describe environ- ments,suchas"abusapproaching"[8].Open-vocabulary supportisaddedbyextensionssuchasObjectFinder[19].
- Multimodal Frameworks: For complete situational awareness, unified pipelines combine OCR, detection, and captioning.
- Business Systems: While OrCam MyEye [11]offers partial edge inference, apps such as Seeing AI [9] and Lookout [10] offer cloud-based services. User-centered trade-offs are highlighted by comparative studies [17].

TABLEI
COMPARISONOFCOMMERCIALASSISTIVESYSTEMS

System	Platform	Connectivity	KeyFeatures	OfflineSup-port
SeeingAI	MobileApp	Cloud	Object&textrecognition, scenedescription	No
Google Lookout	MobileApp	Cloud	Text reading,currency recog-nition,productlabels	No
OrCamMy-Eye	Wearable	PartialEdge	OCR,facerecognition,real-timeaudiofeedback	Yes(partial)

Beyond academic prototypes, commercial deployments of- fer useful insights. Although cloud-based apps like Lookout and Seeing AI achieve high accuracy, they rely on connec- tivity, which raises privacy and latency issues [9],[10]. Inlow-connectivityenvironments,wearableslikeOrCamMyEyeprovideimprovedusabilityandfasteredgeinference[11].Allthingsconsidered,the taxonomyandcomparisonshow the variety of solutions as well as the practical compromises betweeninclusivity,accuracy,responsiveness,andportability.

III. MODELS AND ALGORITHMS

Assistivetechnologiesrelyheavilyondeeplearningmodels, which are generally divided into three categories: text recog- nition, object detection/scene understanding, and image captioning. Each makes it possible for systems to identify, decipher, and convey visual information: object detection guarantees safe navigation [2],[3], OCR facilitates reading [13],[14], and captioning offers semantic context [8].

These models are integrated into unifiedpipelines thatlink sensors, inference modules, and user feedback instead of being used separately. For example, a wearable can read bus numbers (OCR), identify cars (object detection), and describe theenvironment(captioning).Suchintegrationisdemonstrated by commercial tools such as Seeing AI [9], Google Lookout [10], and OrCam MyEye [11], while usability trade-offs are highlighted by user studies (e.g., NavWear [18], OrCam vs. Seeing AI [17]).

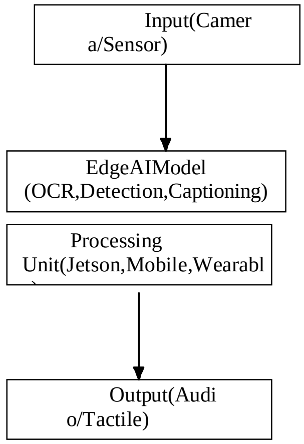

Fig.3.Compactassistivepipeline:inputisprocessedbyedgeAImodels(YOLO [2], SSD [3], OCR [13],[14]) to produce feedback.

A. ImageCaptioning

From visual input, captioning models produce descriptions in natural language. By aligning text with visual features, transformer-based methods perform better than RNNs, and multimodal pretraining (e.g., depth-based captioning [8]) en- hances fluency. High latency and wearable computing costs are obstacles. Model distillation, pruning, and quantizationare some of the solutions. Prototypes based on Jetson achieve 5–10 frames per second [16], which is slower than cloud but sufficient for static descriptions.

B. ObjectDetectionandSceneUnderstanding

Real-time pedestrian and obstacle detection is essential for navigation.

- YOLO:Oftenutilizedinnavigationprototypes,itstrikes a balance between speed and accuracy (about 30 frames per second on embedded GPUs) [2].
- SSD: Mobile-friendly and lightweight, but less precise [3]. Faster R-CNN: High precision, but too sluggish for deployment on the edge.

By simulating spatial arrangements, scene interpretation is further enhanced (e.g., car + pedestrian in crosswalk). Withthe majority of training data coming from urban settings inthe West, dataset bias restricts generalization [15]. The goalof synthetic data and domain adaptation is to lessen this.

TABLEII
EVALUATIONOFWELL-KNOWNOBJECTDETECTIONMODELSFOR ASSISTIVEAPPLICATIONS

Model	FPS(EdgeDevice)	Latency(ms)	mAP/Accuracy	Notes
YOLOv3/v4	25–30	∼35–40	High(0.55–0.60mAP)	Widelyused;excellentspeed-accuracybalance
YOLOv5(Nano/Tiny)	30+	<30	Medium(0.45–0.55mAP)	Designedforembedded devices,lightweight
SSD(MobileNetbackbone)	20–25	∼40–50	Medium(0.40–0.50mAP)	Portable,effective,slightlylessaccuratethanYOLO
FasterR-CNN	<10	>100	High(0.65–0.70mAP)	Veryaccuratebutslow; unsuitableforreal-timeassistiveuse

EfficientDet (D0/D1)	15–20	~50–60	High(0.60–0.65mAP)	Efficientscaling,butrequiresmorecompute

C. TextRecognition (OCR)

OCR makes it possible to read documents, currency, and signs. CNN feature extraction, RNNs/Transformers, and at- tention for robustness under noise are all combined in modern methods [13],[14]. The primary obstacle to multilingualOCR is still that complex scripts (such as Arabic, Hindi, and Chinese) continue to produce high error rates, despite Latin scripts performing well. Multilingual pretraining and effective architectures like MobileOCR, which are appropriate for edge devices, are examples of solutions.

D. Discussion

For more comprehensive feedback, multimodal assistive systems combine detection, OCR, and captioning. Despite advancements, significant obstacles still exist:

- Latency/energy:Complexmodelsputembeddeddevices under stress, necessitating optimization and compression [16].
- Datasetbias:Culturalandgeographicdiversityislacking in training data [15].
- Multilingual OCR: Non-Latin scripts continue to have low accuracy [13],[14]. The user's assessment: Blind participants are rarely included in studies, with the ex- ceptionsofOrCamvs.SeeingAI[17]andNavWear[18].

Futuredevelopmentsincludelightweighttransformersfor on-device deployment, federated learning for privacy- preserving adaptation, and hybrid edge–cloud inference.

IV. EDGE AI HARDWARE PLATFORMS

Because they allow for offline inference, low latency, and user privacy, edge AI platforms are essential to assistive systems [16]. However, there are trade-offs between cost, portability, power, and throughput when choosing hardware [16].

The NVIDIA Jetson Series The de facto standard for deep learningattheedgeistheJetsonfamily:-Nano:15–20FPS, □60 ms latency, appropriate for basic navigation and OCR. - TX2:30+framespersecondwithmoderatepower,commonly utilizedinwearableprototypes.-OrinNano:60+FPSat<20 ms, effective for multimodal V-L models [16].

On Nano, lightweight YOLO versions facilitate obstacle detection [2], and Orin makes it possible to integrate real-time captioning and OCR.

TABLEIII
BenchmarkSummaryofEdgeAIPlatforms

Device	FPS	Latency(ms)	Power(W)	Notes
JetsonNano	15–20	~60	5–10	Low-cost;basicOCR/detection
JetsonTX2	30–35	~30	7–15	Balanced;wearable-friendly
JetsonOrinNano	60+	<20	10–20	Runsmultimodal/transformers
RaspberryPi+CoralTPU	25–30	~35	5–7	Budget;task-specificaccel.

A. OtherPlatforms

TheRaspberryPiwithCoralTPUisadesirableoption forlow-costOCRordetectionbecauseitprovides25–30 FPS at 5–7 W. In contrast to Jetson devices, its task-specific acceleration restricts flexibility.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

B. BenchmarkMetrics

Performanceisassessedusing:-Throughput:$T=\frac{N_{\text{frames}}}{t_{\text{total}}}$-Latency:$L=\frac{1}{T}$-EnergyEfficiency:$E=\frac{N_{\text{frames}}}{P\cdot t_{\text{total}}}$
Fairplatformcomparisonsaremadepossiblebythese metrics [16].

C. Discussion

Real-timemultimodalinferenceismadepossiblebysophis- ticated devices like Orin Nano, but they come with a price tag and a thermal load. Although lighter platforms (Coral TPU, Nano) are more affordable and energy-efficient, they can only be used for specific tasks.
Importanttrade-offs:Powerversusaccuracy:Orinusesmore power but can handle larger models. Cost versus flexibility: Although less versatile, Nano and Coral are more affordable.
Portability versus performance: TX2 strikes a balance be- tween wearable integration and throughput.
Hardware must therefore be matched to the requirementsof the application: Orin for multimodal captioning, TX2 for navigation, and Nano/Coral for OCR.

V. CHALLENGES

The broad use of AI-driven assistive technologies is ham- pered by a number of obstacles, despite significant advance- ments. Trade-offs between accuracy and latency, dataset bias, multilingualconstraints,energyefficiency,andreproducibility are important concerns. In order to guarantee reliable, in- clusive, and useful real-world systems, these issues must be resolved.

A. Accuracyvs.Latency

Accuracy and real-time performance must be balanced in devices with limited resources. While lightweight models (MobileNet, Tiny-YOLO) are quicker but less accurate, high- accuracymodels(suchastransformersandlargeCNNs)have higher latency [2],[3]. While OCR can withstand delays for accuracy, navigation aids prioritize low latency, even with modest accuracy, for safety. This trade-off is best illustratedby wearable Jetson Nano prototypes [16],[18]. Pruning, quantization, and knowledge distillation are some solutions.

B. DatasetBias

Generalization is limited because benchmarks frequently favor Western urban environments [15]. OCR trained on En- glish scripts does not work well on Arabic, Hindi, or Chinese text, and navigation models trained on Cityscapes perform poorly in rural or non-Western environments. This discrep- ancy is demonstrated by commercial apps such as Google Lookout [10]. Open-vocabulary systems like ObjectFinder, diverse datasets, domain adaptation, and synthetic data are all examples of mitigation [19].

C. MultilingualSupport

Text-to-speech and OCR systems need to be able to handle a variety of scripts and styles. Character Error Rate (CER) reveals gaps in inclusivity, with high error rates in non-Latin scripts [13],[14]. For example, OrCam MyEye [11] only sup- ports Hebrew and English. For privacy-preserving adaptation, promisingmethodsincludetransformer-basedOCR,federated learning, and multilingual pretraining.

D. EnergyEfficiency

Wearable technology is subject to stringent power budgets. $E_{\text{total}}=P\cdot t_{\text{inference}}$istheenergyconsumption,where$P$ is the power draw. Optimizing battery life is essential for continuoususe(suchasnavigationaids)[16].Effectivetactics include accelerators (Coral TPU), compression (quantization, pruning), and adaptive inference (e.g., lower frame rate when stationary).

TABLEIV
KEYCHALLENGESANDPOSSIBLESOLUTIONSINASSISTIVEAI

Challenge	PotentialSolutions
Accuracyvs.Latency	Modelpruning,quantization,knowledge distillation,andlightweightarchitectures(MobileNet,Tiny-

	YOLO)[2],[3],[16],[18]
DatasetBias	Open-vocabulary recognition (ObjectFinder), geographically diversedatasets, synthetic data, and domainadaptation [10],[15],[19]
MultilingualSupport	Transformer-basedOCR,federatedlearn-ing, multilingual support, and multilingualpretraining for privacy-preserving adapta-tion [13],[14]
EnergyEfficiency	Task-specificaccelerators(CoralTPU), hardware-awarecompression,andadaptiveinference techniques [16]
Reproducibility	Open-sourceimplementations,transparent evaluationprotocols,andstandardizedbenchmarks [16],[18]

E. Reproducibility

Validation is hampered by a lack of standardized bench- marksandopen-sourcecode[16].Forafaircomparison,com- mon datasets and edge hardware benchmarks are necessary. TransparencyisemphasizedbyinitiativeslikeNVIDIA'sEdge AI Benchmarking suite [16] and fairness projects like Gender Shades [15]. The significance of standardized evaluation is further highlighted by deployments such as NavWear [18].

F. Discussion

Inconclusion,theprimaryobstaclesare:

- Accuracy vs. latency: refined through distillation, quan- tization, and pruning [2],[3],[16],[18].
- Datasetbias:mitigatedbydomainadaptationanda variety of datasets [10],[15],[19].
- Multilinguallimits:addressedwithmultilingualpre- training and wider benchmarks [13],[14].
- Energy efficiency: enhanced through effective hardware and adaptive inference [16].
- Reproducibility:callsforopen-sourceimplementations and standardized protocols [16],[18].

OrCamMyEye[11],GoogleLookout[10],andSeeingAI
[9] are examples of commercial systems that demonstrate the capabilities and constraints of existing approaches. Building scalable, equitable, and user-centered assistive AI requires overcoming these obstacles.

VI. FUTURE DIRECTIONS

A number of exciting research directions are made possible bythequickdevelopmentofassistivetechnologyfortheblind and VI. These approaches tackle the main issues of latency, dataset bias, multilingual support, and real-world usability found in existing systems, in addition to increasing user interaction.

A. Voice-basedInteraction

Accessibility can be greatly increased by using speech asthe main modality. Multilingual support and real-time tran- scription are made possible by large-scale automatic speech recognition (ASR) models like Whisper.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

This enables con- versational interaction where users can ask questions about theirenvironmentandgetcontext-awareresponseswhencom- bined with vision-language frameworks. Early prototypes that combine speech and vision modules have shown how this approach improves usability, especially in environments witha variety of cultures and languages where visual-only systems are inadequate [9],[18].

B. DepthSensingand3DPerception

Byenablingprecise3Denvironmentalmapping,depthsens- ing technologies—such as LiDAR, stereo cameras, and low-costdepthsensors—offersafernavigationbeyond2Dvision.It has been demonstrated that combining depth information with monocular estimation enhances obstacle detection in dynamic and cluttered environments [4]. Real-world navigation proto-typeslikeNavWearconfirmthebenefitofmultimodalsensing, lowering errors compared to image-only methods [18].

C. Bias-awareandFederatedLearning

A significant drawback is still dataset bias, particularly in non-Western contexts. By training models directly on user de-viceswithoutexchangingrawdata,federatedlearningprovides a privacy-preserving solution. In addition, bias-aware loss functions can reduce contextual and demographic differences, ensuring consistent performance across diverse populations [15]. These approaches, which guarantee inclusivity while protecting user privacy, have already been tried in healthcare AI and have great potential for accessibility applications. SystemslikeObjectFinderalsohighlighthowopen-vocabulary recognition can expand inclusivity by moving beyond fixed datasets [16],[19].

D. Cloud–EdgeHybridFrameworks

The computational power of cloud servers and the low- latency advantages of edge computing are combined in hybrid architectures.Whilethecloudcanbeusedforcomputationally demanding tasks, edge devices ensure responsiveness and offlineavailability.Thisequilibriumcanbestatedasfollows:

$$T\text{hybrid}=\alpha T\text{edge}+(1-\alpha)T\text{cloud}, \quad (1)$$

wherethedivision of computation is controlled by $\alpha\in[0,1]$. Future systems can further optimize αdynamically to balance accuracy, efficiency, and energy use [12],[16]. Commercial applicationslikeSeeingAIalreadyemployahybridsetup[9].

E. Outlook

When combined, these patterns point to the multimodal, bias-aware,andhybriddesignoftheupcominggeneration of assistive technology. Federated learning will guarantee inclusivity, depth sensing will allow for a richer environ- mental awareness, voice-based interaction will offer a natural communicationchannel,andhybridframeworkswillstrike a balance between usability and performance. By moving in these directions, assistive AI will progress from prototypes to reliable,practicalsolutionsthatarebothtechnologicallysound and socially inclusive.

VII. CONCLUSION

Thissurveycollectedstudiesonhardwareplatforms,system architectures, and models of AI-powered assistive technolo-giesfortheblindandvisuallyimpaired.Usingthetimeline ofevolution(Fig.1),taxonomyofsystemtypes(Fig.2), and generic processing pipeline (Fig. 3), we charted the progression from simple sensing modules to contemporary multimodal, edge-enabled systems. Comparisons of hardware platforms (Table III) and commercial applications (Table I) also brought to light practical trade-offs between accuracy, latency, energy consumption, and usability.

Advances in deep learning, such as transformers, convo- lutional networks, and multimodal pretraining, have improved OCR[13],[14],objectdetection[2],[3],andimagecaptioning [8]. Deployments on embedded platforms like Jetson Nano, TX2, and Orin Nano demonstrate the growing feasibility of **edgeAI**forlow-latency,privacy-preservingassistivesystems [12],[16]. However,enduringproblemspersist,including limitedreproducibility[16],stringentwearableenergybudgets[12], multilingual constraints [13],[14], and dataset bias [15].

All things considered, AI-driven accessibility is developing quickly but is still in its infancy. The community can hasten thedevelopmentofscalable,dependable,andsociallyequitableassistivetechnologiesbytacklingopenchallengesandembracinginclusive,usercentereddesign.Asasurveypaper,thisworkprovidesresearchersandpractitionerscreatingthenextgenerationofvisuallyimpairedsystems withareferenceandaroadmap.

REFERENCES

[1] WorldHealthOrganization,"Worldreportonvision,"WorldHealthOrganization,Geneva,Switzerland,2019.[Online].Available:https://www.who.int/publications/i/item/9789241516570

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only lookonce:Unified,real-timeobjectdetection,"inProc.IEEEConf.ComputerVisionandPatternRecognition(CVPR),pp.779–788,2016.[Online]. Available:https://ieeexplore.ieee.org/document/7780460

[3] W.Liu,D.Anguelov,D.Erhan,C.Szegedy,andS.Reed,"SSD:Single shot multibox detector," in Proc. European Conf. ComputerVision (ECCV), pp. 21–37, 2016. [Online]. Available: https://arxiv.org/abs/1512.02325

[4] K. Lee and M. Tanaka, "Blind navigation assistance for visually im-paired based on local depth hypothesis from a single image," Sensors,vol. 21, no. 4, pp. 1453–1465, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877705813016214

[5] A. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices forvisually impaired: A survey," Sensors, vol. 20, no. 11, pp. 1–28, 2020.[Online]. Available: https://www.mdpi.com/1424-8220/20/11/3202

[6] L. Chen, H. Wu, and J. Zhao, "An effective obstacle detection systemusing deep learning advantages to aid blind and visually impairednavigation,"IEEEAccess,vol.10,pp.44210–44222,2022.[Online].

[7] Available:https://ieeexplore.ieee.org/document/7514991

[8] A.Alghamdi,A.A.Mutlag,andR.AbdAlhameed,"AdeeplearningbasedassistivesystemforthevisuallyimpairedusingYOLOobjectdetection,"IEEEAccess,vol.9,pp.114–124,2021.[Online]. Available https://scholar.google.com/scholar?hl=en&assdt=0%2C5&q=A.+Alghamdi%2C+A.+A.+Mutlag%2C+and+R.+Abd-Alhameed%2C+%E2%80%9CA+deep+learningbased+assistive+system+for+the+visually+impaired+using+YOLO+object+detection%2C%E2%80%9D+%5Cemph%7BIEEE+Access%7D%2C+vol.+9%2C+pp.+114%E2%80%93124%2C+2021.+%5BOnline%5D.&btnG=

[9] Y.HuangandC.Li,"Animagecaptioningmodelbasedonbidirectionaldepthresidualsanditsapplication,"MultimediaToolsandApplications,vol.81,pp.23245–23260,2022.[Online].Available:https://dl.acm.org/doi/abs/10.1007/s10462-023-10488-2

[10] Microsoft,"SeeingAI,"Microsoft,2023.[Online].Available:https://www.microsoft.com/en-us/ai/seeing-ai

[11] Google,"GoogleLookout,"Google,2023.[Online].Available:https://play.google.com/store/apps/details?id=com.google.android.apps.accessibility.reveal

[12] OrCam,"OrCamMyEye,"OrCamTechnologies,2023.[Online].Avail-able: https://www.orcam.com/en/myeye/

[13] NVIDIA Corporation, "NVIDIA Jetson AGX Orin Series ProductBrief," NVIDIA, 2022. [Online]. Available: https://developer.nvidia.com/embedded/jetson-modules

[14] R. Agyeman and B. Rinner, "Decentralized text-based person re-identification in multi-camera networks," IEEE Access, vol. 12, pp.172125–172148, 2024. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10756694

[15] B.Shi,X.Bai,andC.Yao,"Anend-to-endtrainableneuralnetworkfor image-based sequence recognition and its application to scene textrecognition,"IEEETrans.PatternAnal.Mach.Intell.,vol.39,no.11,pp. 2298–2304, 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7801919

[16] T. Buolamwini and G. Gebru, "Gender Shades: Intersectional AccuracyDisparities in Commercial Gender Classification," in Proc. Conf. Fair-ness, Accountability, and Transparency (FAT), pp. 77–91, 2018. [On-line]. Available: https://proceedings.mlr.press/v81/buolamwini18a.html

[17] D. Minott, S. Siddiqui, and R. J. Haddad, "Benchmarking Edge AIPlatforms:PerformanceAnalysisofNVIDIAJetsonandRaspberryPi5 with Coral TPU," in Proc. IEEE SoutheastCon, pp. 1384–1389, 2025.[Online]. Available: https://ieeexplore.ieee.org/document/10971592

[18] K.Vashistha,P.Mandal,andA.Sinha,"ComparativeusabilitystudyofOrCamMyEyeandMicrosoftSeeingAIforvisuallyimpairedusers,"DisabilityandRehabilitation:AssistiveTechnology,vol.18,no.1,pp.1–12,2023.[Online].Available:https://www.researchgate.net/publication/353519592EvaluationandComparisonofArtificialIntelligenceVisionAidsOrcamMyEye1andSeeingAI

[19] M. Shilkrot, R. Gonzalez-Franco, and P. Maes, "NavWear: A wear-able navigation aid for the visually impaired with real-world userevaluation," in Proc. ACM CHI Conf. Human Factors in ComputingSystems, pp. 1–12, 2021. [Online]. Available: https://dl.acm.org/doi/10.1145/3411764.3445342

[20] H. Li, Y. Wang, and Z. Luo, "ObjectFinder: Open-vocabulary assistiveobject recognition for the blind," in Proc. IEEE Conf. Computer Visionand Pattern Recognition (CVPR), pp. 11562–11571, 2022. [Online].Available: https://ui.adsabs.harvard.edu/abs/2024arXiv241203118L/abstract