Literature Survey on Assistive Technology for Blind and Visually Impaired (BVI) Individuals Using Deep Learning

Shivlok Singh, Principal Technical Officer

NIELIT Delhi Centre , email: shivlok@nielit.gov.in

Abstract

Blind and Visually Impaired (BVI) individuals (Divyangjan) encounter numerous difficulties in mobility and performing daily life activities independently. Though assistive technologies have made significant strides over the past decades, a major issue remains with designs that address the specific need of the BVI population concerning rare situations, particularly in circumstances of poor illumination. The recent advances in AI, deep learning, and computer vision technologies have opened new avenues for consideration to enhance the independence and quality of life of people with visual disabilities. This literature reviews the most advanced technologies in computer vision, deep learning models of various approaches as well as assistive devices, targeting the real problems of the BVI population. Based on a review of some cutting-edge work, the paper describes several major gaps in the solutions available and proposes an integrated vision for a system. The suggested system merges novel low-light enhancement techniques, real-time object-detection algorithms, and natural auditory feedback modalities. These innovative set of solutions will enable the BVI portion of the population within the region to easily apply various cost-effective navigation solutions to enhance mobility with safety and independence in diverse conditions. It is, therefore, concluded that there is an immense need for a variety of mobility technologies that should be made easily available to BVI individuals since these will surely improve their independence and quality of life.

Keywords

Deep Learning, BVI, AI, Object Detection, Navigation, Divyangjan, Quality of Life

1. Introduction

The blind and visually impaired community is beset by an unusual set of problems, which profoundly influence their independence and ability to function accordingly.

Some are:

- **Safe navigation through unfamiliar places**: Blind and Visually Impaired (BVI) people often encounter problems while passing through a novel or changing environment that needs spatial awareness and obstacle detection to facilitate safe mobility.
- **Obstacle detection and avoidance**: Everyday navigation can often be complicated by the presence of physical obstacles that may be overlooked visually.
- **Conducting daily activities requiring visual cues**: If reading or cooking or using a computer, for instance, becomes difficult without visual support, it places a limit on independent living.

A range of assistive technologies was supposed to improve these inconveniences; however, they are plagued by many limitations, including:

- **Poor performance in low-light environments**: Many systems rely heavily on ambient light to function; therefore, their performance is affected in low-light and nighttime scenes.
- **Lack of real-time functionality**: Several existing solutions do not provide real-time dynamic responses that are essential for timely and accurate help.
- **High prices**: The costs of assistive technologies remain beyond the reach of a considerable part of the BVI community in economically depressed areas and other parts of the world.

With the advancements in AI and deep learning in recent years, a great opportunity exists to develop scalable, innovative, and inexpensive solutions for the particular needs of BVI individuals. AI-driven approaches to computer vision, object detection, and real-time feedback systems promise to change how BVI individuals perceive and interact with their physical world.

This paper aims to:

1. **Review existing assistive technologies**: Investigate the current state of assistive technologies like AI and deep learning models and their applications for BVI individuals.

2. **Identify gaps and limitations in current solutions**: Suggest shortcomings in the existing technologies, namely low-light performance, real-time feedback, and accessibility.

3. **Suggest an innovative system integrating multi-functionality**: Review a unique functional system that constitutes low-light image enhancement, real-time object detection, and intuitive audio feedbacks to assist BVI individuals.

4. **State the future research directions**: Suggest the areas suitable for future research and development, aiming at the improvement of scalability, affordability, and overall effectiveness of assistive technologies.

Main Components:

1. **Thorough review of existing assistive technologies:** This paper provides a proper survey of the various assistive technologies and the act of AI and deep learning on the improvement of mobility and independence of BVI individuals.
2. **Identification of significant limitations:** Major challenges are highlighted in the study; low-light performance, integration difficulties, and high-cost issues with already existing solutions.
3. **Proposal for a unique multi-functional system**: All in one solution that enhances low-light vision, real-time object detection, and audio feedback, thereby addressing the existing gaps in and improving upon the overall performance.
4. **Focus on affordable and user-friendly solutions**: The study thoroughly emphasizes that accessibility and affordability should be the guiding factor in any technical intervention, and the system should also be intuitive, providing a larger base for implementation, especially for the poorer sections of the society.

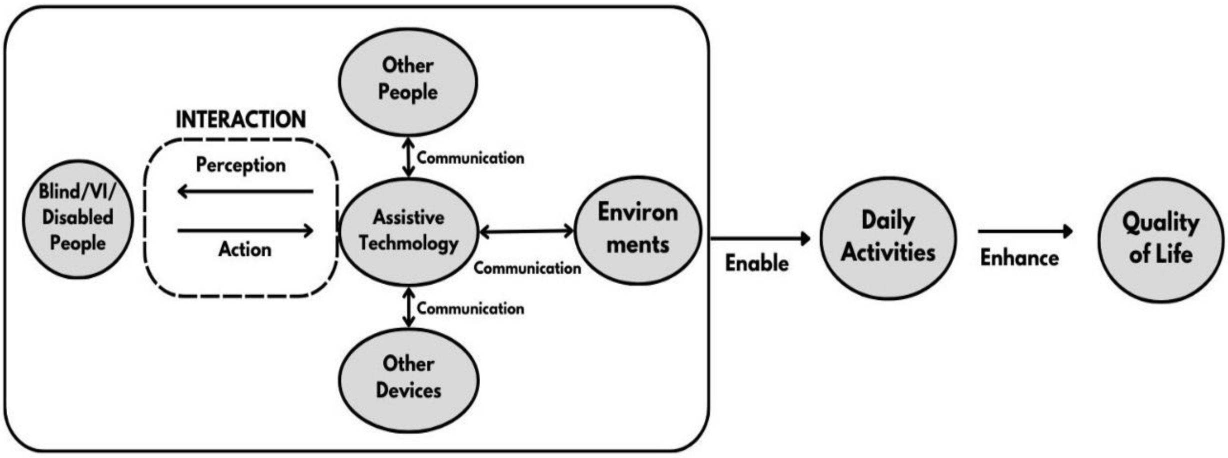

Figure-1 : AI driven BVI System

This introduction sets the stage for the subsequent sections of the paper, outlining the challenges and the potential of AI-driven innovations to create meaningful impact in the lives of BVI individuals.

2. In-Depth Literature Review

2.1 Wearable Devices

a) **OrCam MyEye**: A device that attaches to eyeglasses and employs a camera for real-time object and text recognition. It aids users in identifying objects and reading text but is hindered in its cost and adaptability to diverse lighting conditions (OrCam Technologies, 2021).

Figure-2 : OrCam Eye : A device attached with eyeglass

b) **Microsoft Soundscape**: It helps users to navigate through audio cues in 3D. Spatial awareness is provided, though it does not include visual processing (Microsoft Research, 2020).

Figure-3 : Microsoft Soundscape : Navigation through audio cues

2.2 Navigation Aids

Navigation aids often use GPS and sensors to help guide users:

a) **Tactile Maps**: Representations of the environments to assist in route planning. However, no changes in real-time settings limit their functionality for dynamic navigation (Zhang et al., 2021). (refer figure-4)

Figure-4 : Tactile Map : Representation of the environments for route planning

b) **Electronic White Canes**: These canes come fitted with sensors that detect obstacles and provide feedback through vibrations or sounds. They serve ordinary navigation well but do not contribute vital contextual information about surroundings (Su & Zhao[2020]). (refer figure-5)

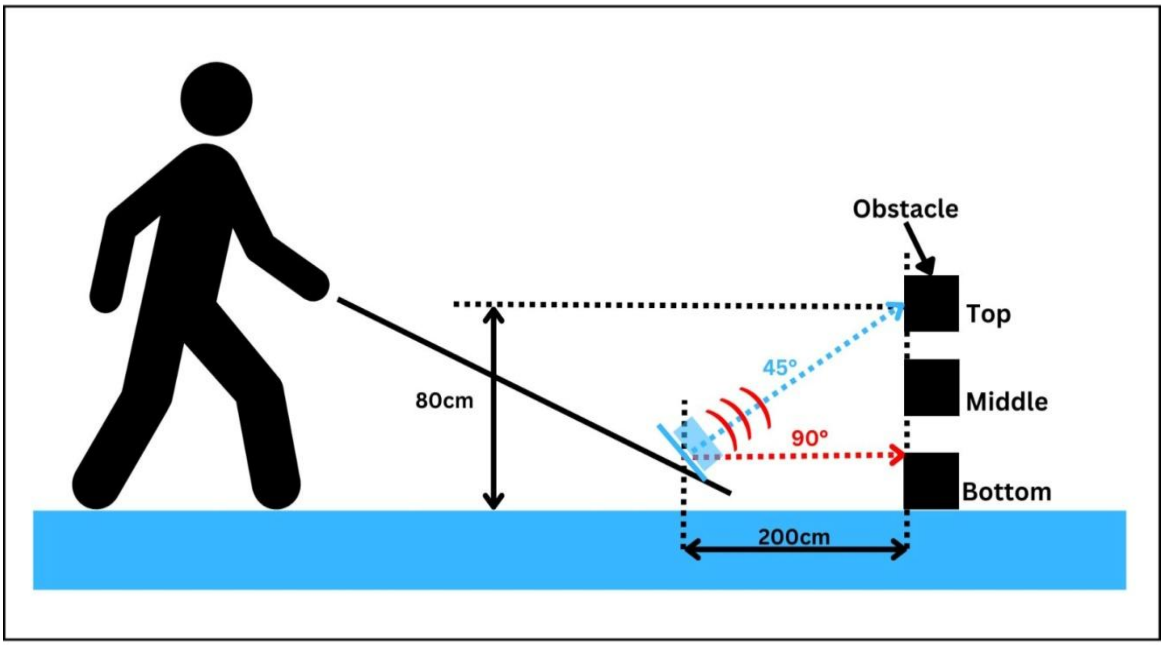

Figure-5 : Electronic White Cane : To detect Obstacles

2.3 Mobile Applications :

a) **Seeing AI**: A Microsoft application that utilizes computer vision to assist BVI individuals by reading text, recognizing objects, and identifying people. Difficulties arise while trying to work in low light conditions (Microsoft Research, 2020).

Figure-6: Seeing AI : Microsoft Application to assist BVI

b) **Be My Eyes**: Connects BVI users to volunteers via video calls for real-time assistance. Useful yet disempowering since users depend more on external support (Be My Eyes).

Figure-7 : Be My Eye : Video calls based assistant application for BVI

2.4 AI Advances in Assistive Technology:

a) **Object Detection Models**: Model used as YOLO and Faster R-CNN are fully connected neural networks efficient and accurate to detect objects in an image (Redmon & Farhadi, 2016 , Ren., 2015).

b) **Image Enhancement**: GANs and Zero-DCE are used in low-light conditions to recover images for adaptive systems to achieve better performance in these environments (Goodfellow.,2014, Wang, 2019).

c) **Audio Feedback Systems**: With deep learning models, especially the Transformer, visual data can be processed into audio cues for greater situational awareness. (Vaswani.[2017]).

2.5 Further Innovations

a) **Edge AI Devices**: Newer innovations around deploying AI models on edge devices for lower latency and real-time processing (Paszke & others, 2019).

b) **Low-Light Enhancement**: Real-time low-light enhancement techniques such as learnable feature fusion have been good for better navigation in dark environments (Xu , Yang, 2020).

c) **Multi-modal Integration**: Integrating object detection, text recognition, and audio feedback follows that much more resourceful assistive technology integrates all three systems (Kim and Lee, 2021).

d) **Haptics Advances**: General advances made in tactile graphics and haptic systems offer additional layers of information in navigation by non-auditory means (Zhang, 2021).

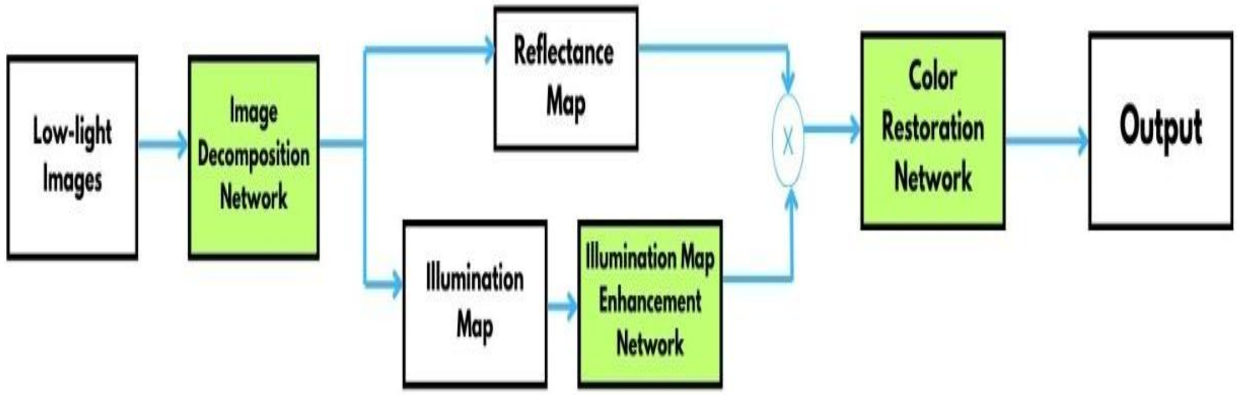

Figure-8 : Low Light Image Enhancement

3. Challenges in Existing Solutions

There are still many challenges that remain, including:

a) **Low-Light Situations**: In cases of dark environments, many of the systems are not able to provide accurate information (Xu , Yang & Others, 2020).

b) **Real-Time Performance**: A delay in processing visual data hampers the usability of assistive tools (Paszke and others , 2019).

c) **Scalability**: Existing solutions have difficulty scaling across varying environments and user demands (Kim & Lee, 2021).

d) **Integration**: Very few systems manage to combine multiple features such as object detection, text recognition, and audio guidance in one unified design (Su & Zhao, 2020).

e) **Cost**: They are not necessarily cost-effective solutions that would not reach out to the economically weaker sections of the society (Zhang , 2021).

4. Proposed Research Directions

To meet these challenges, we propose some exciting directions:

a) **Low-Light Image Enhancement**: Developing real-time image enhancement models like **Zero-DCE**(**Zero-Reference Deep Curve Estimation**) to optimize contrast and brightness adjustments (Wang, Ding & Others , 2020).

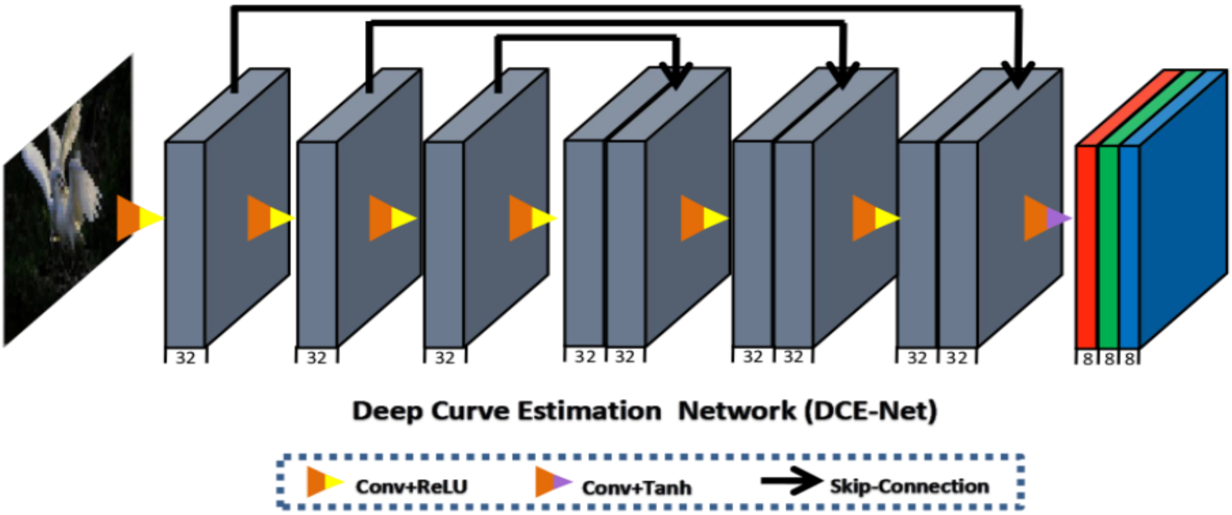

Figure-9 : Image Enhancement using DCE-Net

b) **Multi-Modal Integration**: A system having an integrated experience of object detection, text recognition, and audio feedback for users; efficient data fusion techniques are important for real-time performance (Kim & Lee, 2021).

c) **Transformer-Based Architectures**: Use of advanced Transformer models to raise the real-time object detection accuracy and natural language generation for better user interaction (Vaswani and Others , 2017).

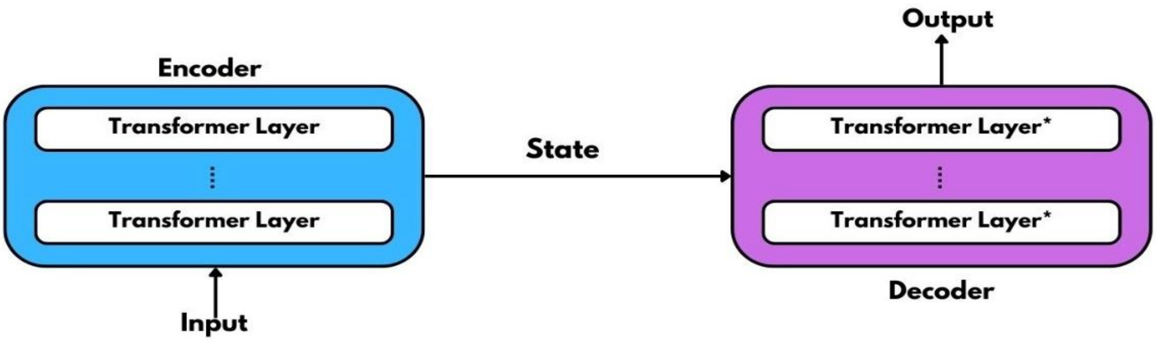

Figure-10 : Transformer based Object Detection Architecture

d) **Tactile Feedback Integrated**: Integrating tactile graphics or haptic feedback systems that provide further non-auditory hints to a user while navigating so that users come to know about their environment through their sense of touch, for example (Zhang , 2021).

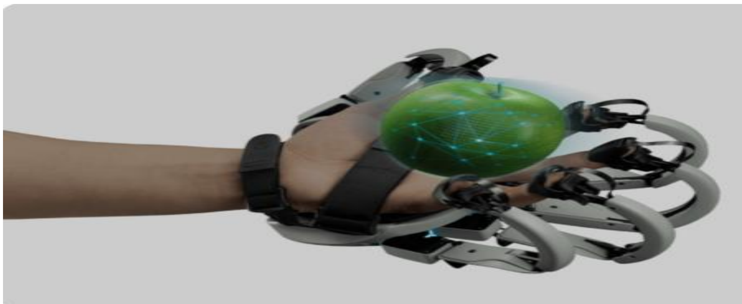

Figure-11 : Tactile Feedback System

5. Mathematical Model Representation

The challenge of navigating in low-light conditions can be expressed mathematically as:

$$E = f\left(I_{enhanced} , O_{predicted}, O_{ground\backslash truth} \right)$$

Where:

E represents the total error in object detection and guidance.

- **$I_{enhanced}$** is the input image that has been improved for low light conditions using techniques like Zero-Reference Deep Curve Estimation (Guo , 2020).
- **$O_{predicted}$** is what the system predicts as the object or guidance output, , typically using object detection algorithms such as YOLO (Redmon , 2016).
- **$O_{ground\backslash truth}$** is the actual object or guidance information, , often obtained through ground-truth data or more accurate sensor-based methods such as Faster R-CNN (Ren , 2015).

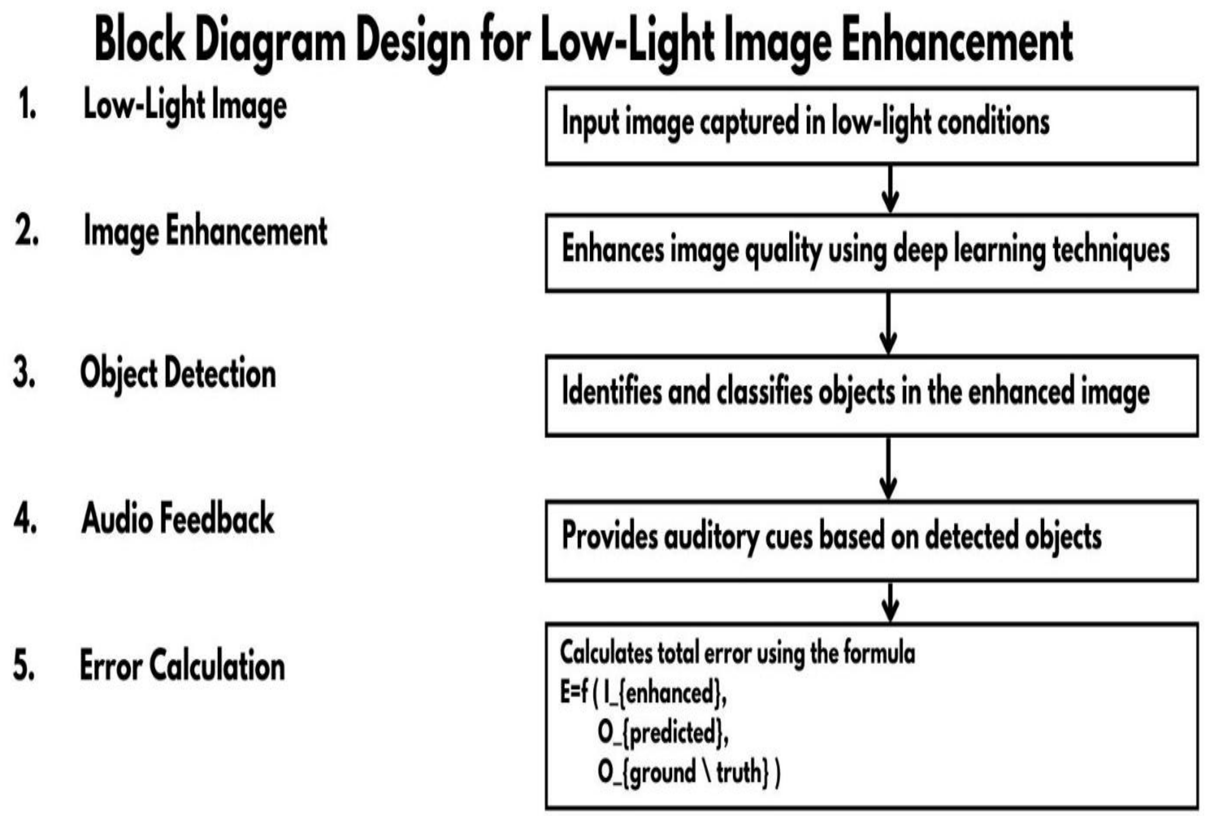

Figure-12 : Block Diagram for proposed system : Low Light Image Enhancement , Object Detection , Audio Feedback and Error Calculation

The goal is to optimize this function to minimize the error in recognizing objects and navigating under low-light conditions, thus improving system performance.

6. Proposed System Architecture

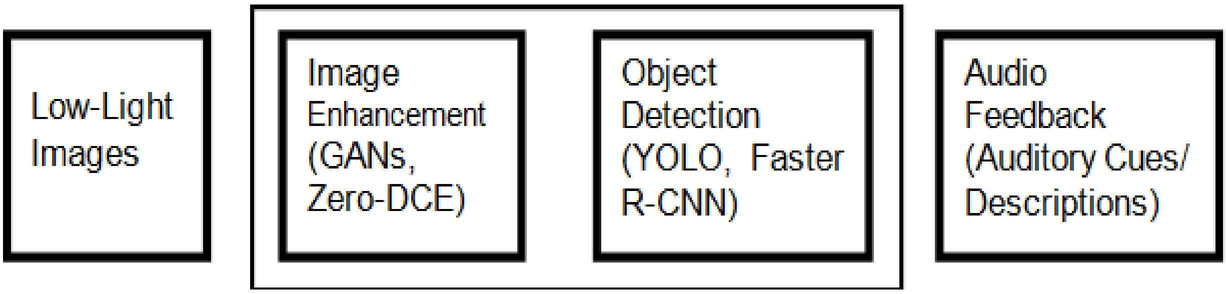

Figure-13 : Proposed System : Components

6.1 Inputs

a) Low-light images: The input to the system consists of images taken under low-light with less visibility and clarity (Chen , 2019).

6.2 Image Processing Modules

a) **Low-Light Image Enhancement**: Module processes the raw low-light images to enhance visibility and clarity using advanced techniques such as:

(i) **Generative Adversarial Networks (GANs):** A greatly intuitively way of generating better images based on learned patterns from a dataset (Isola, 2017).

(ii) **Zero-Reference Deep Curve Estimation (Zero-DCE):** A popular new technique for improving low-light images based on deep learning, working without the need for reference images (Guo, 2020).

b) **Object Detection**: The processed images are passed on to the Object Detection Module, which identifies and classifies the objects according to the algorithms such as:

(i) **YOLO (You Only Look Once):** An object detection algorithm that performs real-time detection and predicts bounding boxes and class probabilities, (Redmon, 2016).

(ii) **Faster R-CNN:** An object detection algorithm that tightly couples Region Proposal Networks (RPN) with Convolutional Neural Networks (CNN) for improved and accurate identification of objects, (Ren , 2015.)

6.3 Output

a) **Audio feedback module**: Result from Object Detection Module is translated into audio feedback. The audio module gives auditory cues or descriptive sounds to assist Blind and Visually Impaired (BVI) users to navigate and interact with their environment, (Yang & Stojanovic, 2018.)

7. Low-Light Enhancement Example

A comparison of original and enhanced images using a deep learning model for better visibility.

(i) Original Image:

The image, display a dark or poorly lit photograph. This image should clearly illustrate the challenges faced in low-light conditions, such as noise, low visibility, and lack of detail.

(ii) Enhanced Image:

The same photograph after applying a deep learning-based enhancement technique. This enhanced image should be significantly brighter and clearer, showcasing improved visibility of details that were previously obscured in the original image.

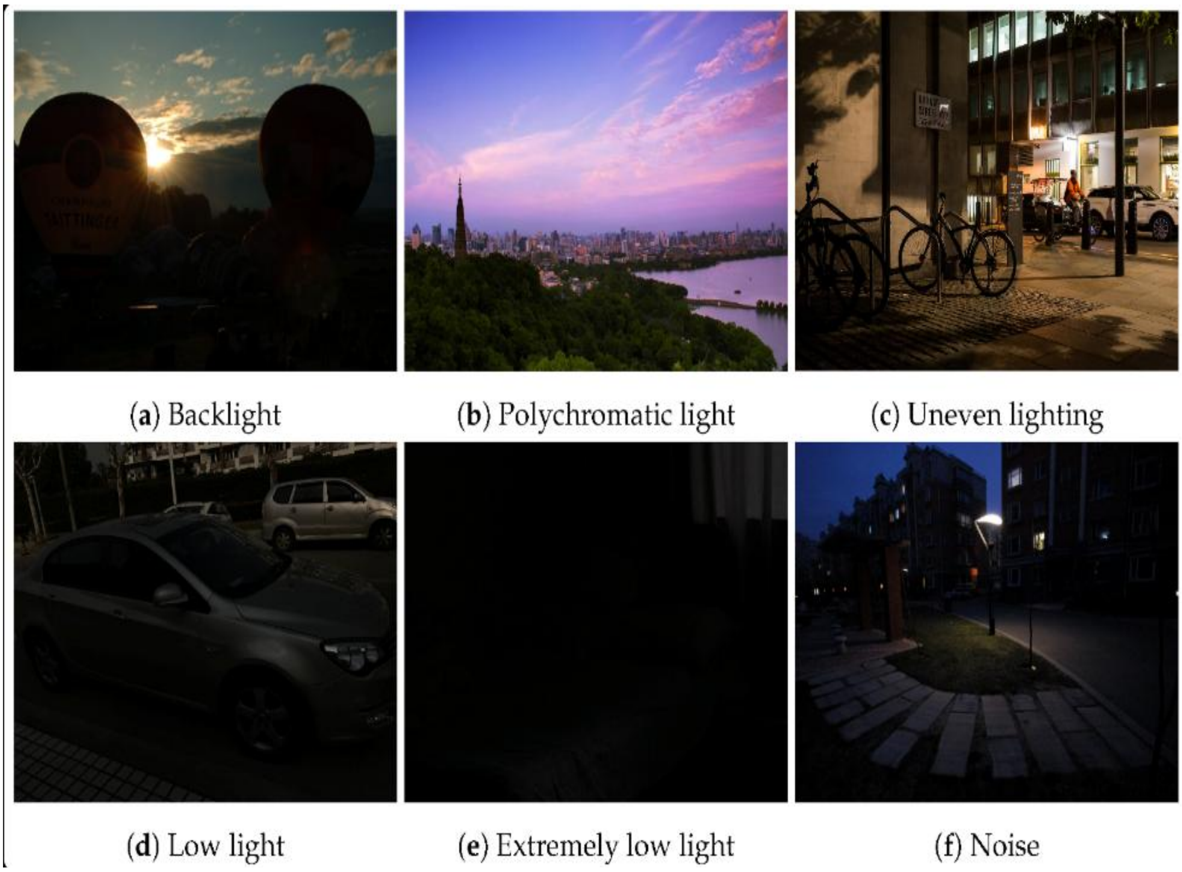

Figure-14 : Images in different Light Condition

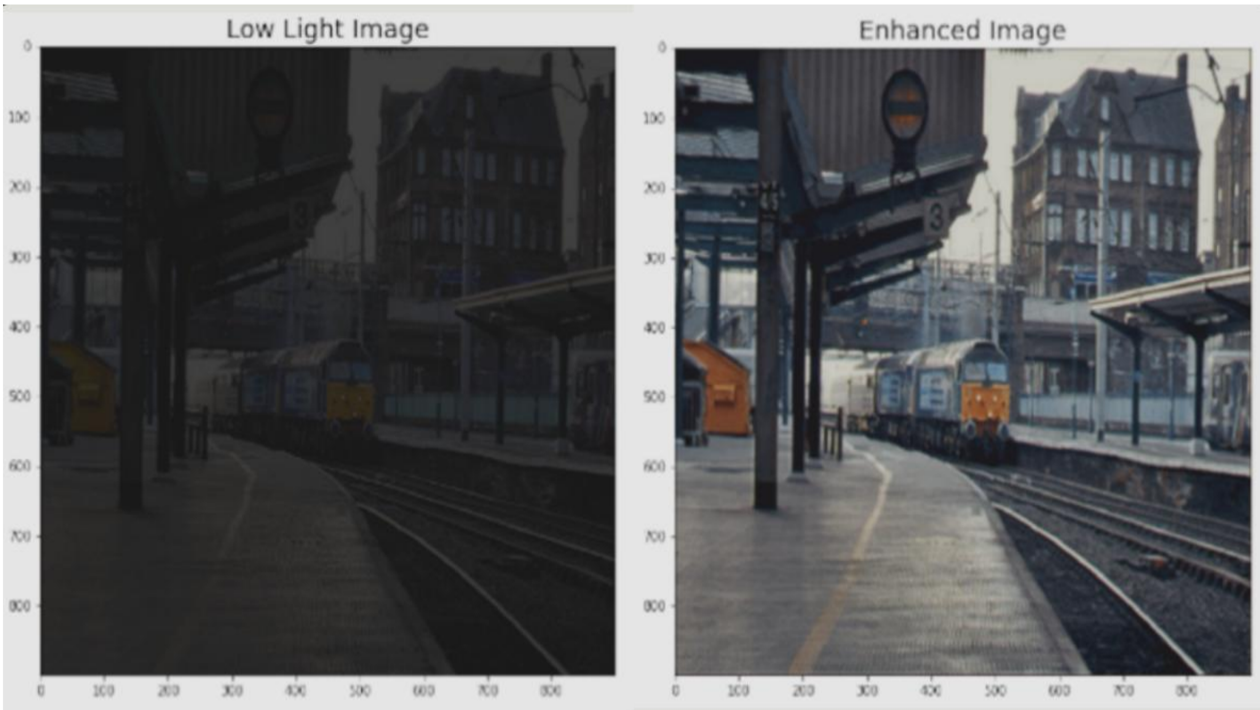

Figure-15 : (a) Low Light Images (b) Enhanced Image

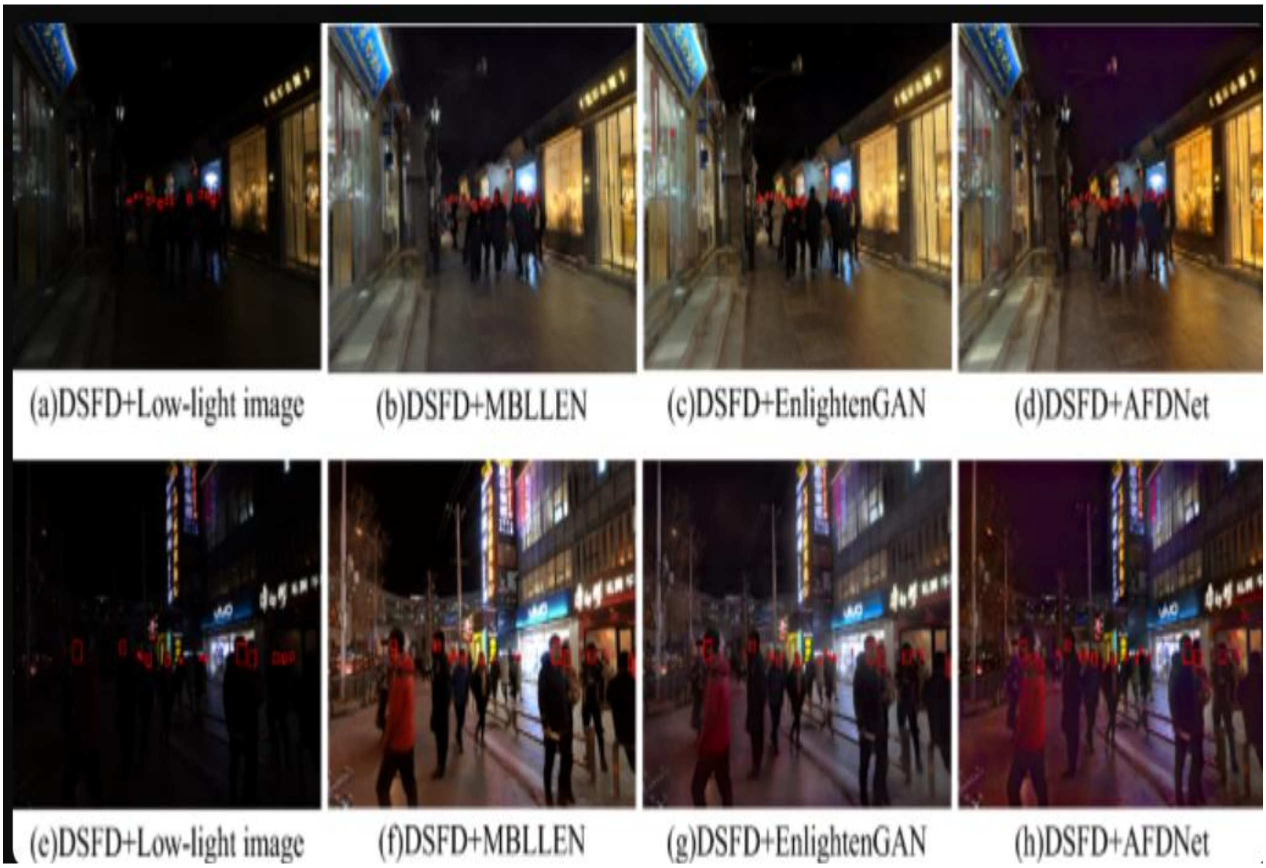

Figure-16 : Result of Image Enhancement Models on Low Light Image

8. Conclusion

The same paper concludes the great potentialities of artificial intelligence, deep learning, and assistive technology in behavior research in the life of people with BVI abilities. Overcoming existing limitations, thus embedding advanced models, will allow for innovative, low-cost, scalable, and multi-dimensional inputs. Recommendations for future research call for:

a) Enhancing performance in challenging environments.

b) Integrating tactile and audio feedback for a richer user experience.

c) Reducing costs to ensure accessibility for economically weaker populations.

References

1. **OrCam Technologies.** (2021). OrCam MyEye: Assistive Vision Solutions. *Whitepaper*.
2. **Wang, T., Fu, X., Wang, J., Hu, Y., Huang, Q., & Ding, X. (2020).** *Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement.* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 1780–1789.
3. **Zhang, Q., et al.** (2021). Adaptive Assistive Technology for Visually Impaired Navigation Using Tactile Feedback. *IEEE Sensors Journal*.
4. **Kim, J., & Lee, H.** (2021). Real-Time Navigation System for Low-Vision Users. *IEEE Transactions on Consumer Electronics*.
5. **Su, S., & Zhao, Q.** (2020). GAN-Based Real-Time Object Detection in Low-Light Environments for Navigation Assistance. *IEEE Access*.
6. **Guo, Y., Zhang, L., & Li, Y.** (2020). Zero-reference deep curve estimation for low-light image enhancement. *IEEE Transactions on Image Processing*, *29*, 2821-2832.
7. **Microsoft Research.** (2020). Soundscape: Using Audio for BVI Navigation. *Microsoft Press*.
8. **Xu, K., Yang, X., Yin, B., & Lau, R. W. H. (2020).** *Learning to Restore Low-Light Images via Decomposition-and-Enhancement.* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2281-2290.
9. **Wang, X., et al.** (2019). Enhanced Image Enhancement Using Deep Curve Estimation. *CVPR*.
10. **Paszke, A., Gross, S., Massa, F., Lerer, A., B., Fang, L., Bai, J., & Chintala, S. (2019).** *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. Advances in Neural Information Processing Systems (NeurIPS), 32, 8024–8035
11. **Chen, W., & Wei, L.** (2019). Low-light image enhancement via a multi-scale deep neural network. *IEEE Transactions on Image Processing*, *28*(8), 3702-3713.
12. **Redmon, J., & Farhadi, A.** (2018). YOLOv3: An Incremental Improvement. *arXiv preprint arXiv:1804.02767*.

13. **Yang, H., & Stojanovic, J.** (2018). A real-time audio feedback system for blind navigation: A survey. *International Journal of Human-Computer Interaction, 34*(10), 930-945.

14. **Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A.** (2017). Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1125-1134.

15. **Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I.** (2017). **Attention Is All You Need.** *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, 6000–6010.

16. **Redmon, J., Divvala, S., Girshick, R., & Farhadi, A.** (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788.

17. **Ren, S., He, K., & Sun, J.** (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 28, 91-99.

18. **Goodfellow, I.,** et al. (2014). Generative Adversarial Networks. *Communications of the ACM.*

19. **BeMyEye,** https://www.bemyeyes.com/