International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 6 ǁ 2022 ǁ PP. 226-231

AI Assistant For Visually Impaired

PAULSON K R, PRIYADHARA PUTHRAN S O, SREEKRISHNA KRISHNAKUMAR, ROBY REJI
Department Of Computer Science
Sahrdaya College of Engineering And Technology, Thrissur, Kerala, India

Abstract—*In today's advanced hi-tech world, the importance of independent living is recognized in the situation of visually impaired people who are socially restricted. They endure in unusual environments with no physical assistance. Because most tasks rely on visual information, visually impaired people are at a disadvantage because vital information about their surroundings is unavailable. With recent advancements in inclusive technology, it is now possible to provide further assistance to those with visual impairments. This project proposes to use Artificial Intelligence, Machine Learning, Image and Text Recognition to assist folks who are blind or visually impaired. The concept is applied in a way that emphasizes voice assistant, picture recognition, obstacle detection, and e-books, among other things. The system can recognize items in the environment and perform text analysis to recognize text in a hard copy document using voice commands. It will be an effective approach for blind individuals to engage with the environment and use technology's features.*

Index Terms— *Visually impaired, image and text recognition, artificial intelligence, voice assistant, obstacle detection*

Date of Submission: 25-05-2022 Date of acceptance: 05-06-2022

I. INTRODUCTION

BLINDNESS is one of the most frequent disabilities inthe world. Over the last few decades, the number of people who have become blind as a result of natural causes oraccidents has increased. People who are partially blind havefoggy vision, only notice shadows, and have poor night visionor tunnel vision. On the other hand, a completely blind personhas no vision. According to World Health Organization data,there are approximately 2.2 billion visually impaired or blindpeople worldwide [1]. Blind individuals have traditionally useda white cane to help them navigate their surroundings, albeitthis method does not give information for moving obstacles ap-proaching from afar. Furthermore, white canes cannot identifyhigher impediments above the knee level. Another method forassisting the blind is to use trained guide dogs. Trained dogs,on the other hand, are expensive and hard to come by. Severaltypes of wearable or hand-held electronic travel aids(ETAs) have been proposed in recent studies [2]-[9]. The majorityof these gadgets include a variety of sensors that map theenvironment and offer voice or sound alarms via headphones.The reliability of these devices is influenced by the quality ofthe real-time auditory signal. Many contemporary ETAs lack areal-time reading aid and have a poor user interface, high cost,limited portability, and no hands-free access. As a result, thesegadgets are not well-liked by the blind, and they need to be improved in terms of design, performance, and dependability for usage in both indoor and outdoor environments.

Many contemporary ETAs lack a real-time reading aid and have a poor user interface, high cost, limited portability, and no hands-free access. As a result, these gadgets are not well- liked by the blind, and they need to be improved in terms of design, performance, and dependability for usage in both indoor and outdoor environments.

In this paper, we suggest a new visual aid method for people who are fully blind. The following are the distinguishing characteristics of the suggested design that characterize itsnovelty:

1) Indoor and outdoor navigation with an integrated readingassistant in a hand held form.
2) With a low-end hardware setup, the core processes based on machine learning algorithms has been implemented with cloud architecture.
3) Real-time video, image and audio analysis for object detection and identification, obstacle identification.

The proposed arrangement can identify both fixed and moving items in real time and provide audio feedback to the visually impaired. Furthermore, the tablet includes an in-built reading aid that can read text from any paper. The planned visual aid system's design, construction, and performance evaluation are discussed in this article.

Section II of this paper covers the available literature on blind navigation aids, emphasizing their advantages and dis- advantages. Section III details the prototype's design and op- eration, whereas Section IV discusses the experimental setup for performance evaluation. Section V presents a summary of the findings based on statistical analysis. Finally, Section VI brings the article to a close.

II. PROBLEM DEFINITION

According to the World Health Organization (WHO), ap- proximately 286 million people are visually impaired and 40 million are blind in 2019. The problem affects blind people on a regular basis. They can't even walk without assistance. Traditional aids such as the white cane and guide dogs are effective in resolving issues. They don't have enoughknowledge to avoid all of the stumbling blocks. They areunable to seek for and pick up any objects they desire.

III. RELATED WORKS

ETAs, electronic orientation aids, and positioning locating devices are the three types of electronic assistance for the visu- ally handicapped. ETAs help you navigate safely by detecting, warning, and avoiding objects[10]-[12]. ETAs work in few steps; sensors are used to collect data from the environment, which are then processed through a computing device to detect an obstacle or object and give the user a feedback corresponding to the identified object. By sending out a 40 kHzsignal and receiving a reflected echo from the object in frontof it, the ultrasonic sensors can detect an object within 300 cm.

The pulse count and time-of-flight are used to determine the distance (TOF). Smart glasses [2], [9] and boots [12] with ultrasonic sensors have already been presented as a visual assistance for the blind. Katzschmann's [13] novel technique employs an array of infrared TOF distance sensors pointingin various directions. To emit and detect IR pulses reflected from barriers, Villanueva and Farcy [14] use a white cane with a near-IR LED and a photodiode, respectively. For the blind, cameras [15, [16] and binocular vision sensors [17] have been employed to gather visual data.

The acquired data is processed using a variety of devices and procedures. The images recorded by the camera wereprocessed using a Raspberry Pi 3 Model B+ and open com- puter vision (OpenCV) software [18]. Google tango [3] and other platforms have also been used. A wearable device is used with cloud-enabled computation [2]. Another alternative for processing the acquired data is to use a field-programmable gate array [19]. The collected photos are preprocessed to eliminate noise and distortion. The Gaussian filter, grey scale conversion, binary image conversion, edge detection, and cropping [20] are all used to manually process the images. TheTesseract optical character recognition (OCR) engine is then used to extract text from the processed image [21]. The stereo picture quality assessment [17] uses an innovative techniqueto choose the best image from a large number of candidates. The best image is then sent into a cloud-based convolutional neural network (CNN) that has been trained on huge data. Most devices provide audio feedback through a headset or a speaker. The audio is either a synthetic voice generated by the text-to-speech synthesis system [22] or a beep sound generatedby a voice user interface [23]. Some systems additionally use vibrations and tactile feedback.

Ando [24] showed out a haptic device that looks like awhite cane but has an incorporated smart sensing methodand an active handle that detects an obstruction and generates vibrations that replicate a real experience on the cane handle. Direct cane is a typical white cane-like system [13] that moves on wheels and uses steering servo motors to guide the wheels by using ultrasonic sensors to detect impediments. The user must always hold the device in their hand in this system, however several systems that give a hands-free experience are widely available. NavGuide [12] and NavCane [25] are assistance devices that identify impediments up to the knee level using numerous sensors. Wet floor sensors are included inboth NavGuide and NavCane. NavCane is a global positioning system (GPS) with a mobile communication module that may be connected into white cane systems.

Xiao [4] shows a context-aware navigation framework that uses GPS to give visual cues and distance sensing as wellas location-context information. With the support of Wi-Fi connectivity through the Internet, the platform may also access geographic information systems, transportation databases, and social media. Lan [26] presented a smart glass system thatcan detect and recognize road signs in cities in real time, such as public restrooms, restaurants, and bus stops. This system is small, portable, and adaptable. However, simply readingthe road signs may not provide enough information for ablind person to feel safe in an outdoor setting. Because public signs differ from city to city, the system will be unable to recognize a sign that has not been recorded in the system's database. Hoang [20] created an assistive system that uses a mobile Kinect and a matrix of electrodes to detect and warn about obstacles. Since the sensors are always situated inside themouth during navigation, the device has a complicated design and an uncomfortable setup. Furthermore, it is more costly andless portable. Islam [27] gave a thorough overview of sensor- based walking aids for the blind and visually challenged. The authors found critical characteristics that make an effective walking assistant. Low-cost, simple, and lightweight design with good inside and outdoor coverage are among them. Several blind user groups, software developers, and engineers provided feedback. In addition, Dakopoulos and Bourbakis [10] Identified 14 structural and operational characteristics thatdefine an optimum ETA for the blind.

Despite repeated efforts, many existing systems do not fully integrate all functionalities and are frequently limited by cost and complexity. Our key contribution was to create an ETA prototype for the blind that was simple, low-cost, portable, and hands-free, with text-to-speech conversion capabilities for basic, everyday indoor and outdoor use. While the suggested system currently lacks advanced functionality such as wet floor detection and climbing stair detection, reading of traffic signs, GPS use, and a mobile communication module, the flexible design allows for future modifications and enhancements.

IV. DESIGN OF THE PROPOSED DEVICE

The hardware component, Raspberry Pi board 3 B+ with Pi cam module is used to capture video input, get the audio input from the user and transmit collected data to another com- ponent of implementation Microsoft cloud services Azure. The program has been written in Python and the packages included OpenCV, Tensorflow. The hardware setup of the proposed device is shown schematically in Fig. 1, while a snapshot of the real device prototype is shown in Fig. 2.

Our work in cloud platform consisted of developing our ownalgorithms for processing the incoming data as well as utilizingservices provided by Microsoft, Azure Cognitive Services. Multiple methods have been used in the object detection including OpenCV. Azure Cognitive Services (from now on it will be referred as ACS) employs multiple algorithms for face detection, OCR, speech to text and text to speech. In addition to that we have developed our own programs for the above mentioned processes and compared the results to evaluate the accuracy. Without any prior knowledge of AI or machine learning, a developer can easily deploy the AI functionality using the ACS which employs different advanced machine algorithms. Simply invoking the ACS API allows a developer to quickly deploy AI functionality. The API may read text from an image, a document, or a URL. Handwritten documentstext (Only English language is supported as of now), Text taken from printed documents, and text taken from mixed- language sources. Using a REST API, the Azure Cognitive Speech service turns text into synthesised speech and receivesa list of available voices for a location.
The Raspberry Pi 3 B plus model was chosen as the

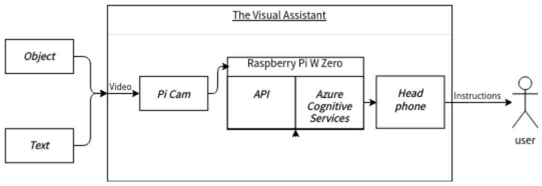

Fig. 1. The suggested system's hardware configuration. The visual assistant receives the image as input, processes it using the Raspberry Pi processor,and provides audible feedback via headphones.

Fig. 2. Model that has been proposed. The Raspberry Pi with the camera attached to the board.

functional device due to its low cost and portability. It alsohas multiprocessing capabilities, unlike many other systems.

A. Implementation

Figure 3 depicts the Raspberry Pi 3 B plus Model's con- nection to the rest of the system's components. The system is designed to help visually impaired people to detect obstacles and read text images. It guides the user with audio output fromspeaker or headphone. The device gets triggered only when it isgiven an audio input through the microphone connected with raspberry Pi. Audio signals from the hardware are taken by our own program and are used to call Azure functions as well as ACS.

Azure functions are a cloud service available with several modules for OCR, object detection. The audio signals taken from user are undergone speech synthesis to deduce the requiredfacility. Azure functions call the corresponding API. When the user gives audio input for object detection or text reading the API activate camera module attached to the device and input is taken in the form of video or image. Image or videois taken to API and processed. If the user requires text reading facility OCR module is used to retrieve text from the image and for obstacles avoidance it uses object detection module to label the objects. Text produced in this stage is either directly converted to audio at Azure functions itself or give it to user through speaker or the text is taken to the hardware and convertsto audio there at Raspberry Pi facilities. If the user asks for

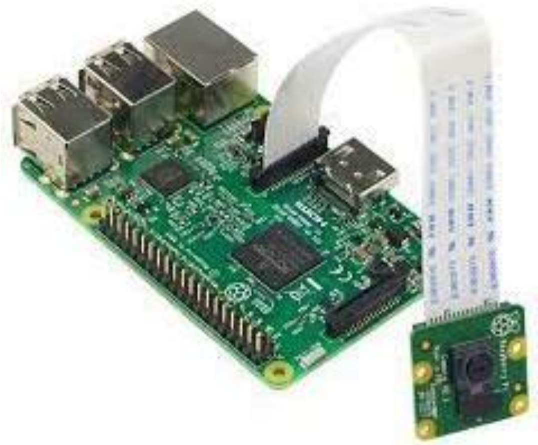

Fig. 3. Raspberry Pi 3 B plus model and related module with cameramodule(basic hardware setup).

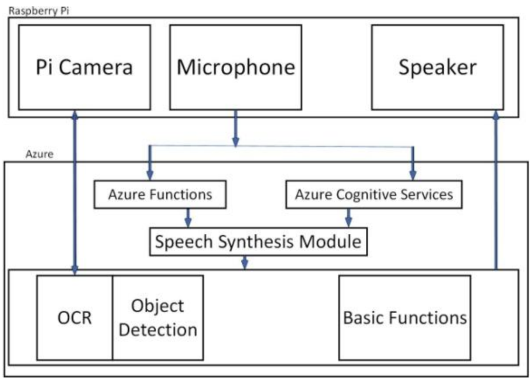

Fig. 4. The suggested system's hardware configuration. The visual assistant receives the image as input, processes it using the Raspberry Pi processor,and provides audible feedback via headphones.

other functions like location or time Azure functions directly call API and give output through speaker.
Azure cognitive Services (ACS) is used to get more accurateresults. The audio input from user is taken by Azure Cognitive Services and undergone speech synthesis. ACS provided with inbuilt functions do the works more accurately and give instructions. Output from Azure functions is compared with ACS. Figure 4 depicts the architecture of the system implemented.

The reading functioning principle is depicted in Figure 6. Without disrupting the object detection process, a picture is acquired from the live video feed. Azure API will extract the texts from the image and save them in a temporary text filein the background. The text from the text file is then spoken out using the text to speech service. The accuracy of OCRis affected by ambient lighting and background, and it works best with a white background and highly lit areas.

V. SYSTEM EVALUATION AND EXPERIMENTS

A. *Evaluation of Object Detection*

The detection of a single object from a video stream is shown in Figure 6. Despite the fact that the background fills the white cane received good marks. The used pre trained model could be retrained with more objects to improve perfor-mance. Under well lit conditions, the reading aide performed admirably. Users pointed out that one of the reading assistant'ssignificant flaws was that it couldn't read texts with tables andgraphics.

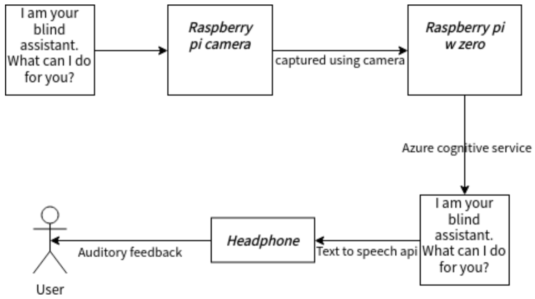

Fig. 5. The reading assistant's workflow. The camera module sends a single frame to the Raspberry Pi, which it processes through Azure. The audio outputis then transformed from the test output.

the majority of the image, the model is able to filter out other bounding boxes and recognize the required object in the framewith 97 percent accuracy.

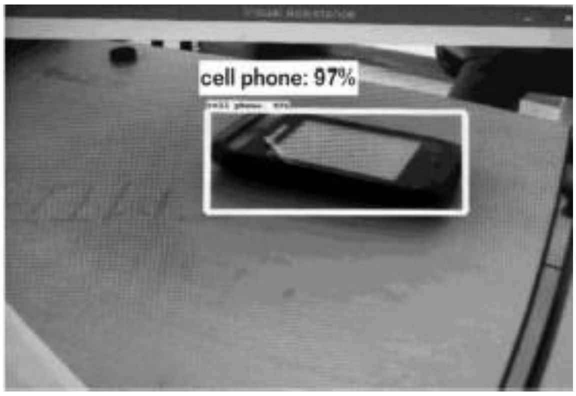

Fig. 6. Detection of a single object. The cell phone can be detected with 97percent certainty using the object detection method.

B. Evaluation of Reading Assistant

Our prototype's built-in reading assistance is put to the test in a variety of lighting circumstances with variable text sizes, fonts, colors, and backgrounds. In a brighter atmosphere, the system performs better since it can easily extract text from thecollected image. When comparing text with different colored backgrounds, it was discovered that the reading assistanceperforms better with a well-illuminated background. Whenthe font color is black, the device operated well in bright and even slightly dark environments, but it failed to read the entire sentence in the dark. The reading assistant had no problems with the green-colored text in a highly illuminated area, but struggled in slightly dim and dark conditions.

VI. RESULTS AND DISCUSSION

Figure 5 depicts the principle of operation of the reading helper. Figure 6 depicts the detection of a single object from a video stream. The comfort level was slightly affected because we only used a prototype to conduct the experiments. However, the proposed device's mobility and preference over the white cane received good marks. The used pre trained model could be retrained with more objects to improve perfor- mance. Under well lit conditions, the reading aide performed admirably. Users pointed out that one of the reading assistant's significant flaws was that it couldn't read texts with tables and graphics

VII. CONCLUSION

This research paper proposes a revolutionary visual aid system for the absolutely blind in the form of a handheld device. The following are some of the proposed device's primary features.

1) For indoor and outdoor navigation, the hands-free, wear- able, low-power, low-cost, and compact design is ideal.
2) The Raspberry Pi w zero Model's low-end computing capacity is used to process complex algorithms.
3) A built-in reading assistance with image-to-text conver- sion features that allows the blind to read text from any document.

The software and hardware features of the proposed blind aide have been thoroughly discussed. The system may also be built and tested in a more complicated outdoor setting thanks to powerful machine learning algorithms and an upgraded userinterface.

REFERENCES

[1]. Blindness and vision impairment, World Health Organization, Geneva, Switzerland, Oct. 2019. [Online]. Available: https://www.who.int/newsroom/ fact-sheets/detail/blindness-and-visual-impairment
[2]. J. Bai, S. Lian, Z. Liu, K Wang, and D. Liu, "Virtual-blind-road following based wearable navigation device for blind people," IEEE Trans. Consum. Electron., vol. 64, no. 1, pp. 136–143, Feb. 2018.
[3]. B. Li et al., "Vision-based mobile indoor assistive navigation aid for blind people," IEEE Trans. Mobile Compute., vol. 18, no. 3, pp. 702–714, Mar. 2019.
[4]. J. Xiao, S. L. Joseph, X. Zhang, B. Li, X. Li, and J. Zhang, "An assistive navigation framework for the visually impaired," IEEE Trans.Human- Mach. Syst., vol. 45, no. 5, pp. 635–640, Oct. 2015.
[5]. A. Karmel, A. Sharma, M. Pandya, and D. Garg, "IoT based assistive device for deaf, dumb and blind people," Procedia Comput. Sci., vol. 165, pp. 259–269, Nov. 2019.
[6]. C. Ye and X. Qian, "3-D object recognition of a robotic navigationaid for the visually impaired," IEEE Trans. Neural Syst. Rehabil. Eng.,vol. 26, no. 2, pp. 441–450, Feb. 2018.
[7]. Y. Liu, N. R. B. Stiles, and M. Meister, "Augmented reality powersa cognitive assistant for the blind," eLife, vol. 7, Nov. 2018, Art. no. e37841.
[8]. A. Adebiyi et al., "Assessment of feedback modalities for wearable visual aids in blind mobility," PLoS One, vol. 12, no. 2, Feb. 2017,Art. no. e0170531.
[9]. J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu, "Smart guiding glasses forvisually impaired people in indoor environment," IEEE Trans. Consum.Electron. vol. 63, no. 3, pp. 258–266, Aug. 2017.
[10]. D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," IEEE Trans. Syst. Man, Cybern. Part C Appl. Rev., vol. 40, no. 1, pp. 25–35, Jan. 2010.
[11]. E. E. Pissaloux, R. Velazquez, and F. Maingreaud, "A new framework for cognitive mobility of visually impaired users in using tactile de- vice," IEEE Trans. Human-Mach. Syst., vol. 47, no. 6, pp. 1040–1051, Dec. 2017.
[12]. K. Patil, Q. Jawadwala, and F. C. Shu, "Design and construction of electronic aid for visually impaired people," IEEE Trans. Human-Mach.Syst., vol. 48, no. 2, pp. 172–182, Apr. 2018.
[13]. R. K. Katzschmann, B. Araki, and D. Rus, "Safe local navigationfor visually impaired users with a time-of-flight and haptic feedback device," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 26, no. 3, pp. 583–593, Mar. 2018.
[14]. J. Villanueva and R. Farcy, "Optical device indicating a safe free path toblind people," IEEE Trans. Instrum. Meas., vol. 61, no. 1, pp. 170–177,Jan. 2012.
[15]. X. Yang, S. Yuan, and Y. Tian, "Assistive clothing pattern recognition for visually impaired people," IEEE Trans. Human-Mach. Syst., vol. 44, no. 2, pp. 234–243, Apr. 2014.
[16]. S. L. Joseph et al., "Being aware of the world: Toward using social media to support the blind with navigation," IEEE Trans. Human-Mach.Syst., vol. 45, no. 3, pp. 399–405, Jun. 2015.
[17]. B. Jiang, J. Yang, Z. Lv, and H. Song, "Wearable vision assistance system based on binocular sensors for visually impaired users," IEEE Internet Things J., vol. 6, no. 2, pp. 1375–1383, Apr. 2019.
[18]. L. Tepelea, I. Buciu, C. Grava, I. Gavrilut, and A. Gacsadi, "A vision module for visually impaired people by using raspberry PI platform,"in Proc.15th Int. Conf. Eng. Modern Electr. Syst. (EMES), Oradea, Romania, 2019, pp. 209–212.
[19]. L. Dunai, G. Peris-Fajarnés, E. Lluna, and B. Defez, "Sensory naviga- tion device for blind people," J. Navig., vol. 66, no. 3, pp. 349–362,
[20]. May 2013
[21]. V.-N. Hoang, T.-H. Nguyen, T.-L. Le, T.-H. Tran, T.-P. Vuong, and N. Vuillerme, "Obstacle detection and warning system for visually impaired people based on electrode matrix and mobile kinect," VietnamJ. Comput. Sci., vol. 4, no. 2, pp. 71–83, Jul. 2016.
[22]. C. I. Patel, A. Patel, and D. Patel, "Optical character recognition by open source OCR tool Tesseract: A case study," Int. J. Comput. Appl., vol. 55, no. 10, pp. 50–56, Oct. 2012.
[23]. A. Chalamandaris, S. Karabetsos, P. Tsiakoulis, and S. Raptis, "A unit selection text-to-speech synthesis system optimized for use with screenreaders," IEEE Trans. Consum. Electron., vol. 56, no. 3, pp. 1890–1897,Aug. 2010.
[24]. R. Keefer, Y. Liu, and N. Bourbakis, "The development and evaluation of an eyes-free interaction model for mobile reading devices," IEEE Trans. Human-Mach. Syst., vol. 43, no. 1, pp. 76–91, Jan. 2013.
[25]. R. Keefer, Y. Liu, and N. Bourbakis, "The development and evaluation of an eyes-free interaction model for mobile reading devices," IEEE Trans. Human-Mach. Syst., vol. 43, no. 1, pp. 76–91, Jan. 2013.
[26]. V. V. Meshram, K. Patil, V. A. Meshram, and F. C. Shu, "An astute assistive device for mobility and object recognition for visually impaired people," IEEE Trans. Human-Mach. Syst., vol. 49, no. 5, pp. 449–460, Oct. 2019.
[27]. F. Lan, G. Zhai, and W. Lin, "Lightweight smart glass system with audio aid for visually impaired people," in Proc. IEEE Region 10 Conf.,Macao, China, 2015, pp. 1–4.
[28]. M. M. Islam, M. S. Sadi, K. Z. Zamli, and M. M. Ahmed, "Developingwalking assistants for visually impaired people: A review," IEEE Sens.J., vol. 19, no. 8, pp. 2814–2828, Apr. 2019.
[29]. T-Y Lin et al., "Microsoft COCO: Common objects in context," Feb. 2015. [Online]. Available: https://arxiv.org/abs/1405.0312
[30]. J. Han et al., "Representing and retrieving video shots in human-centricbrain imaging space," IEEE Trans. Image Process., vol. 22, no. 7, pp. 2723–2736, Jul. 2013.
[31]. J. Han, K. N. Ngan, M. Li, and H. J. Zhang, "Unsupervised extraction of visual attention objects in color images," IEEE Trans. Circuits Syst. Video Technol., vol. 16, no. 1, pp. 141–145, Jan. 2006.
[32]. D. Zhang, D. Meng, and J. Han, "Co-saliency detection via a self- paced multiple-instance learning framework," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 5, pp. 865–878, May 2017.
[33]. G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convo- lutional neural networks for object detection in VHR optical remote sensing images," IEEE Trans. Geosci. Remote Sens., vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
[34]. Y. Yang, Q. Zhang, P.Wang, X. Hu, and N.Wu, "Moving object detection for dynamic background scenes based on spatiotemporal model," Adv. Multimedia, vol. 2017, Jun. 2017, Art. no. 5179013.
[35]. Q. Xie, O. Remil,Y. Guo,M.Wang,M.Wei, and J.Wang, "Object de- tection and tracking under occlusion for object-level RGB-D video segmentation," IEEE Trans. Multimedia, vol. 20, no. 3, pp. 580–592, Mar. 2018.
[36]. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans.Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.