Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)

Volume7-Issue2, January 2024
Paper: 31

Utilizing the text image and a clever, intelligent AIgadget to solve obstacles

Dr. Krishna Dharavathu Ph. D[1], G. Sai Ganesh Reddy[2], K.S.G.S Deekshith[3],

A. Lakshman[4], K. Mounika[5]

[1] Faculty, and [2,3,4,5] Students
Dept. of Electronics and Communication Engineering, PSCMR College of Engineering & Technology, Vijayawada, India.

krishnadharavath4u@gmail.com, ganeshreddy7721@gmail.com,

karedladeeks5@gmail.com, lakshmanarja416@gmail.com,

lakshmimounika1022@gmail.com,

Abstract: *Recent years have seen tremendous progress in the field of image processing, driven by the combination of deep learning, computer vision, and machine learning methodologies. This paper explores the core components of image processing, namely object detection, text detection, and image segmentation. Using convolutional neural networks (CNNs) and deep learning architectures such as R-CNN, object detection is the process of identifying and localizing objects in images. Applications like optical character recognition (OCR) and document analysis are made easier by text detection algorithms, which find and identify text within images. By dividing images into logical sections, image segmentation algorithms allow for accurate item identification and scene comprehension. These techniques find applications in a variety of disciplines, including as autonomous driving, medical imaging, and surveillance. They do this by combining conventional methods with cutting-edge deep learning methodologies. This article underscores the pivotal role of image processing in modern technology and its continuous evolution through ongoing research and innovation.*

Keywords: *optical character recognition (OCR), convolutional neural networks (CNNs), Machine learning (ML), styling.*

1. INTRODUCTION:

As we know, technology has changed significantly over the years and will continue to do so, improving many aspects of life and raising standards of living. The improvement of human civilization has been greatly aided by technological advancement. Technology offers creative approaches to tasks using clever methods. All field's requirements are now more efficient and demand less time and effort thanks to technology [1]. The study and application of various methods for modifying and analyzing digital images is known as image processing. AI-powered text detection is a major development in computer vision that allows computers to recognize and extract text with amazing efficiency and accuracy from photos.

Applications for this technology can be found in many different domains, including augmented reality, automated data entry, document processing, and accessibility for the blind [2]. The object detection module gives the virtual assistant the ability to identify and categorize items in pictures or video feeds. It makes use of deep learning models to

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 31
locate objects in a visual scene, recognize their categories, and detect objects: convolutional neural networks (CNNs) and region-based convolutional neural networks (R-CNNs) [3]. With the use of this feature, the assistant can recognize items in photos, help users identify objects, and carry out actions based on objects it has discovered. In many different fields, including robotics, medical imaging, remote sensing, and entertainment, it is essential [4]. The fundamental goal of image processing is to enhance, restore, or extract information from images in order to get insightful knowledge or improve their quality. Images are transformed by computational techniques and algorithms, such as feature extraction, segmentation, and filtering [5]. Tasks like picture compression, pattern recognition, text detection and object detection are made possible by these processes. Image processing finds use in surveillance systems for security and in medical diagnosis, where it helps spot malignancies [6]. Additionally, it serves as the foundation for developments in artificial intelligence, especially in computer vision applications like facial recognition and autonomous driving. Advances in hardware, software, and algorithms in image processing provide for new possibilities in the analysis and interpretation of visual data as technology progresses [7].

2. LITERATURE SURVEY:

This section describes the earlier designed image processing models such as google lens. A literature review for an image processing entail looking at previously conducted studies, academic papers, and pertinent books that cover a range of topics related to image processing algorithms. The purpose of this survey is to gather baseline data, identify knowledge gaps, and provide guidance for the creation and enhancement of image processing.

K. Wang et al. [8] developed a character detection technique for optical character recognition systems using color clustering, black adjacency graph analysis, and aligning-and-merging analysis, proving effective in experimental studies on 325 640 x 480 pixels images. Wang et al. [9] developed a robust word detection method for images and video frames, achieving a detection rate of 93.9% for English and 92.4% for Chinese text. Chen et al. [10] developed a novel technique for identifying text in complex images and videos, using segmentation, OCR algorithms, localization, machine learning, and text verification. K. Jung et al. [11] conducted a comprehensive evaluation of various text extraction methods from photographs and videos, focusing on text size, style, alignment, orientation, low image contrast, and complex backdrops.

Senevirathne et al. [12] developed a chat program using convolutional neural networks to identify images of old Sri Lankan artifacts, enhancing knowledge exchange and user experience through a user-friendly chatbot. Pardasani et al. [13] developed a method using Python computer vision to recognize hand motions in images for Sign Language, enabling effective communication for mute individuals. Wookjae Maeng et al. [14] developed a chatbot to address Image-based sexual abuse (IBSA), providing emotional support and information, and found it superior to internet search in organization, accessibility, and conciseness.

Nishant et al. [15] proposed a virtual shopping assistant for e-commerce, utilizing image recognition to provide relevant responses to user queries, addressing issues with product features like color or pattern, amidst social media influence. T. Pitchayagan et al. [16] developed a LINE Bot System for diagnosing rice diseases from field images, using

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 31

deep learning neural networks and YOLOv3 deployment model. The system detects diseases in real-time, requiring only 2-3 seconds for detection.

Lavanya et al. [17] utilized image processing to determine if participants were donning masks during the COVID-19 pandemic. They used TensorFlow, Keras library, OpenCV, and a Convolutional Neural Network (CNN) model to analyze a dataset of mask-free and mask-free images captured using a PC's webcam.

Gagandeep Singh et al. [18] utilized Image Recognition and Voice Recognition in their Android application to assist visually impaired individuals in independent living, enhancing their interaction with their surroundings. Akalanka et al. [19] developed a tool that uses machine learning to predict a patient's risk level and analyzes chest CT scan and X-ray images, achieving high accuracy rates of 92% and 96% for CT scan classification and health history data analysis.

Rohit et al. [20] developed a real-time system for image identification, classification, and position estimation, transforming the visible world into an auditory one. The system uses motion, sharpening, and blurring filters, enabling visually impaired people to perceive their environment through voice output. B. Cvetkoska et al. [21] proposed a Smart eHealth Mirror utilizes face recognition technology to enhance posture, identify health issues, and recommend preventive healthcare. The PAA directs users to the correct location of the smart mirror, and image processing allows for precise identification of health issues.

S. Nazim et al. [22] study employs a CNN model and computer vision to recognize images, identify faces, and synthesize festival speech, while smart glasses utilize ultrasonic sensors. Mail Automation plays a vital role in this project. The participant's mail IDs are collected in a spreadsheet, and the mail IDs are inputted into the mail automation command. This command sends the particular certificate of that participant to his/her mail ID and also stores a copy of all the documents in a separate folder in PDF format. Robotic Process Automation provides more accessibility for users to automate complex processes using software bots.

3. METHODOLOGY:

A. Word Detection

The Optical Character Recognition (OCR) tool is all you need to detect text in an image. The Tesseract is one of the greatest OCR systems available for extracting text from images. The tesseract module can be used to integrate Tesseract OCR with Python [11].The following stride dictates the method followed by engine to acquire the text from images uploaded. Preprocessing: Tesseract utilizes preprocessing techniques to enhance the quality of text recognition by converting the input image to grayscale, enhancing contrast [8]. Text Detection: Tesseract analyzes preprocessed images and finds possible text regions by applying techniques such as contour detection and linked component analysis [9]. Character Segmentation: Tesseract is a software that efficiently divides text into individual glyphs or characters by recognizing text regions and determining their placement within these regions [10]. Language Modeling: Based on the features that are gathered, Tesseract employs statistical language models to estimate the most likely character sequence. Language models can increase text recognition accuracy by capturing the likelihood that a given sequence of characters will occur in that language[11].

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 31

B. Face Detection

Getting the Image Ready: To increase face detection accuracy, preprocessing methods could involve shrinking the image, turning it to grayscale, and boosting contrast. Face Detection: To find faces in an image, use a face detection method or a face detection model that has already been trained. i.e. Haar Cascade Classifiers. Face Recognition: You can utilize a pre-trained face recognition model or an algorithm to identify particular people. Feature extraction and comparison with a database of recognized faces are the standard steps in face recognition systems [12].

Labelling the images: The images are labelled according to the user's interest. Model training: The labelled images are made as data set and trained to the model, to recognize the facial features of a particular face and label the person respectively during detection [13]. Display of the Outcome: The detected image possessing the recognized face was displayed in a labelled box.

The steps below specify how the engine will obtain the object from uploaded images.

Data Collection: Assemble a varied dataset with images of the objects you wish to identify. The collection ought to encompass diverse positions, lighting situations, backgrounds, and additional aspects that could impact the recognition of objects [14].

Data Preprocessing: Preprocessing the gathered data will guarantee its quality and uniformity. Resizing images to a consistent size, standardizing pixel values, and enhancing the dataset with methods like rotation, flipping, and noise addition to boost its diversity are examples of preprocessing procedures [15].

Labeling: Add appropriate labels to the photos in the collection that identify the things seen in each image. In supervised learning, the machine learning model gets its knowledge from labeled instances; proper labeling is necessary for this process.

Feature Extraction: Extrapolate significant characteristics from the pictures that will help you differentiate between various objects. Techniques including edge detection, texture analysis, color histograms, and feature descriptors like SURF or SIFT may be used in this step [16].

Model Selection: Select an object identification machine learning model architecture that makes sense. Convolutional Neural Networks (CNNs) are a popular option because of their ability to learn hierarchical features directly from raw pixel data. CNNs have shown exceptional performance in image identification applications [17].

Training: Utilizing the labeled dataset, train the chosen model. By modifying its internal parameters in response to training samples and the given loss function, the model gains the ability to map input photos to the associated object labels [18].

Validation: To evaluate the trained model's generalization performance and spot possible problems like overfitting or underfitting, validate it on a different validation dataset. Based on validation results, modify model architecture and hyperparameters as necessary [19].

Evaluation: Analyze the trained model's accuracy, precision, recall, and other pertinent parameters using a different test dataset. Evaluate the model's performance in comparison to other cutting-edge techniques or baseline methods.

Deployment: Use the learned model to perform practical object identification tasks. This could entail incorporating the model into an embedded system, online service, or application so that it can process incoming images and deliver predictions instantly [20].

Monitoring and Maintenance: Maintain a close eye on the functioning of the deployed model in production, and periodically retrain it with new data to accommodate variations in the object's look, surroundings, or other variables that may impact the accuracy of object recognition [21].

4. FLOWCHART:

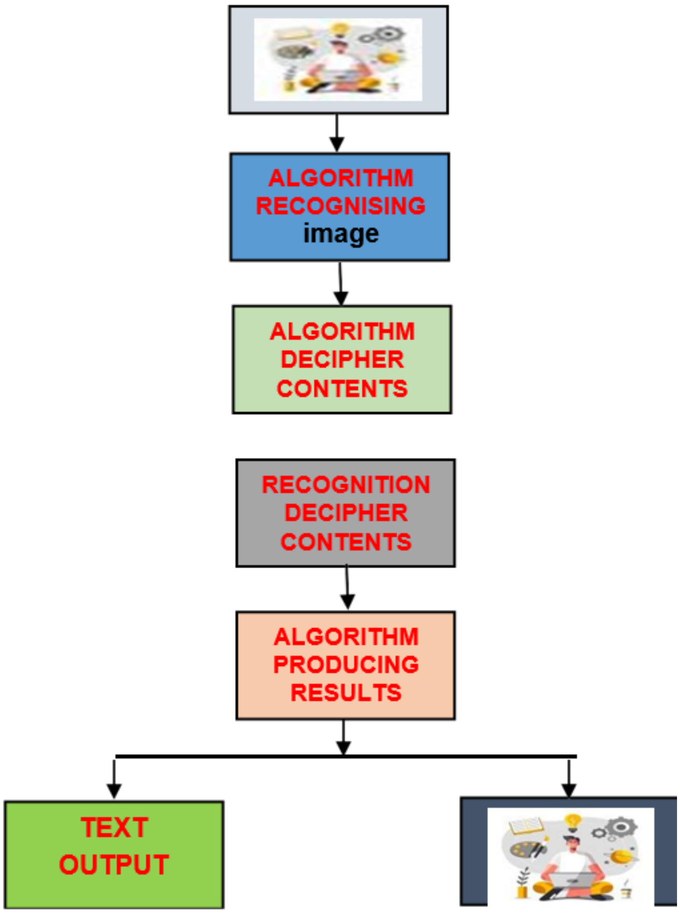

Fig. 1. Image Detection

5. RESULTS AND DISCUSSION:

In this experiment we tried to blend the OCR, facial recognition and object detection into one combination to make it adaptable for the text processing and detection of image elements. We used tesseract, harrassacode to create data set that is to be used to train the model and produce the results. In these results we produced the images of face detection, object detection and word

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 31

decoding from image.

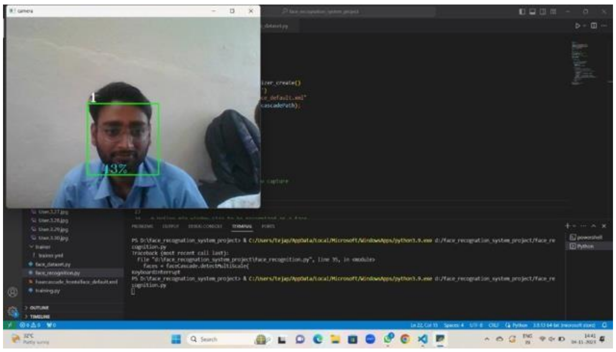

Fig 2: Facial Detection

Fig 3: word detection

Fig 4: Object detection

6. CONCLUSION:

We concluded that image processing tools such as object and text detection are very useful for a wide range of applications. Advances in autonomous driving, document analysis, and monitoring are made possible by their capacity to recognize and locate objects and text inside images. These approaches are still being developed, but they promise increased efficiency and accuracy by utilizing complex algorithms and deep learning techniques. The combination of text and object identification, as technology develops, has enormous potential to transform a variety of sectors and enhance daily chores.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)

Volume7-Issue2, January 2024
Paper: 31

7. REFERENCES:

[1] Shivansh Kohli, U. Verma, Vinay V Kirpalani and R. Srinath, “Dermatobot: An Image Processing Enabled Chatbot for Diagnosis and Tele- remedy of Skin Diseases,” DOI: 10.1109/INCET54531.2022.9824756

[2] CH.M.H. Saibaba, S. waris, S.H. Raju, V. Sarma, V. C. Jadala and Ch. Prasad, “Intelligent Voice Assistant by Using OpenCV Approach,” DOI: 10.1109/ICESC51422.2021.9532956

[3] R S Sai Dinesh, Surendran. R, D Kathirvelan and V Logesh, “Artificial Intelligence based Vision and Voice Assistant,” ICEARS,DOI:10.1109/ICEARS53579.2022.9751819

[4] Charan, Kashyap and Harsha Vardhan Reddy, “Virtual Assistant and Navigation for Visually Impaired using Deep Neural Network and Image Processing,” DOI: https://doi.org/10.21203/rs.3.rs-2867156/v1

[5] Hemlata M. Jadhav, Altaf Mulani and Makarand M. Jadhav, “Design and Development of Chatbot Based on Reinforcement Learning,” https://doi.org/10.1002/9781119861850.ch12

[6] Qixiang Ye and David Doermann, “Text Detection and Recognition in Imagery: A Survey,” IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 37, NO. 7,DOI:10.1109/TPAMI.2014.2366765

[7] R. Lienhart and A. Wernicke, “Localizing and segmenting text in images and videos,” IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 4, pp. 256–268, Apr. 2002.

[8] K. Wang and J. A. Kangas, “Character location in scene images from digital camera,” Pattern Recognit., vol. 36, no. 10, pp. 2287–2299, https://doi.org/10.1016/S0031-3203(03)00082-7

[9] Q. Ye, W. Wang, W. Gao, and W. Zeng, “A robust text detection algorithm in images and video frames,” in Proc. Joint Conf. Inf., Commun. Signal Process. Pac. Rim Conf. Multimedia, 2003, pp. 802–806

[10] D. Chen, J. M. Odobez, and H. Bourlard, “Text detection and recognition in images and video frames,” Pattern Recognit., vol. 37, no. 3, pp. 596–608, 2004.https://doi.org/10.1016/j.patcog.2003.06.001

[11] K. Jung, K. I. Kim, and A. K. Jain, “Text information extraction in images and video: A survey,” Pattern Recognit., vol. 37, pp. 977–997, 2004. https://doi.org/10.1016/j.patcog.2003.10.012

[12] Senevirathne and Vidanagama, “IMAGIBOT – An Image Recognition Chatbot for Sri Lankan Ancient Places,” http://ir.kdu.ac.lk/handle/345/3033

[13] A. Pardasani, A. K. Sharma, S. Banerjee, V. Garg and D. S. Roy, “Enhancing the Ability to Communicate by Synthesizing American Sign Language using Image Recognition in A Chatbot for Differently Abled,” (ICRITO), Noida, India, 2018, pp. 529-532, doi:10.1109/ICRITO.2018.8748590.

[14] Wookjae Maeng and J. Lee, “Designing and Evaluating a Chatbot for Survivors of Image-Based Sexual Abuse,” https://doi.org/10.1145/3491102.3517629

[15] Goel, Nishant, “Shopbot: An Image Based Search Application for E- Commerce Domain” (2017). Master's Projects. 516.DOI: https://doi.org/10.31979/etd.r7a5-6dzf

[16] T. Pitchayagan, K. Kiratiratanapruk, A. Kitvimonrat, Wasin and Sujin, “A system for automatic rice disease detection from rice paddy images serviced via a Chatbot, https://doi.org/10.1016/j.compag.2021.106156

[17] Lavanya Kesa, Shreya, Kirti and Sandeep, “Chatbot with Facemask Detection Technique,” https://doi.org/10.22214/ijraset.2021.35511

[18] Gagandeep Singh, Omkar, Kevin and Nandini, “A Smart Personal AI Assistant for Visually Impaired People,” IRJET, Volume: 07 Issue: 06

[19] B. A. Akalanka, K. D. A. Senevirathne, M. H. V. Dias, W. A. R. Nimantha and C. M. Silva, “Smart Assistant to Ease the Process of COVID-19 and Pneumonia Detection,” 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference, (IEMCON) doi: 10.1109/IEMCON53756.2021.9623105.

[20] P. Rohit, M. S. Vinay Prasad, S. J. Ranganatha Gowda, D. R. Krishna Raju and I. Quadri, "Image Recognition based SMART AID FOR VISUALLY CHALLENGED PEOPLE," 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 1058-1063, doi: 10.1109/ICCES45898.2019.9002091.

[21] B. Cvetkoska, N. Marina, D. C. Bogatinoska and Z. Mitreski, "Smart mirror E-health assistant — Posture analyze algorithm proposed model for upright posture," IEEE EUROCON 2017 -17th International Conference on Smart Technologies, Ohrid, Macedonia, 2017, pp. 507-512, doi: 10.1109/EUROCON.2017.8011163.

[22] S. Nazim, S. Firdous, S. R. Pillai and V. K. Shukla, "Smart Glasses: A Visual Assistant for the Blind," 2022 International Mobile and Embedded Technology Conference (MECON), Noida, India, 2022, pp. 621-626, doi: 10.1109/MECON53876.2022.9751975.