Multi-Disciplinary Journal
ISSN No- 2581-9879 (Online), 0076-2571 (Print)
www.mahratta.org,editor@mahratta.org

Survey on Applications of Object Identification using Deep Learning

Mali Priyanka Mohan,
M.Sc. Student, TMV
priyapratap123@gmail.com

Deepmala Sharma,
deepa0702@gmail.com
Assistant Professor, TMV_Kharghar

Abstract

Human cerebrum takes not exactly a moment to distinguish the area of item inside the picture just as remember it when it makes sure; however, machine needs time and enormous measure of information to do a similar assignment. Deep neural network dependent on convolution neural network gives high precision and extraordinary outcomes in object discovery and order. To prepare deep neural networks, huge measure of information, for example, (pictures and video recordings) and time is required. As computational expense of PC vision is extremely high, move learning strategy, where a model prepared on one undertaking is reused on another connected assignment, gives better outcomes. Through the survey and investigation of deep learning-based article recognition methods lately, this work incorporates the accompanying parts: spine networks, misfortune capacities and preparing methodologies, traditional item identification designs, multifaceted issues, the datasets and assessment dimensions, applications, and future advancement headings. We trust this survey paper will be useful for specialists in the field of item location. The application discussed here is an object detection for blind peoples.

Keywords: Object detection, Deep Learning, Neural Network, Machine Learning, Convolutional Neural Networks

Introduction:

Object Item recognition has a few relations with object order, semantic division, and example division. The subtleties are shown in Fig. 1. Article identification is a significant space of PC vision and has significant applications in logical examination and useful mechanical creation, for example, face discovery [5], text location [6], passerby recognition [7], logo discovery [8], video discovery [9], vehicle location and clinical picture discovery. Then, acquiring from neural organizations and related learning frameworks, the advancement in these fields will create neural organization calculations, and will likewise enormously affect object location procedures which can be considered as learning frameworks. [10]. In any case, because of enormous varieties in perspectives, postures, impediments, and lighting conditions, it's hard to consummately achieve object discovery with an extra item confinement task. Such a lot of consideration has been pulled into this field as of late [11]. The difficult meaning of item location is to figure out where articles are in each picture (object confinement) and which class each item has a place with (object arrangement). In this way, the

pipeline of customary item discovery models can be basically separated into three phases: instructive area choice, include extraction and arrangement.

Figure 1: Object Detection Output Sample [3]

Artificial Intelligence is getting well known in all enterprises with the primary motivation behind improving income and diminishing expenses; by utilizing Machine learning strategy they robotize and streamline their interaction to tackle testing errands effectively. Face recognition and walker location are firmly identified with conventional item identification and primarily cultivated with multi-scale adaption and multi-highlight combination/boosting timberland, individually. The spotted lines demonstrate that the relating spaces are related with one another under specific conditions. It ought to be seen that the covered areas are broadened. Walker and face pictures have ordinary designs, while general items and scene pictures have more mind-boggling varieties in mathematical constructions and formats. Along these lines, phenomenal significant models are required by various pictures. Profound Learning Techniques are likewise utilized for early suggestion to forestall such clinical sicknesses those emerges with maturing and furthermore wrongdoing specialists to limit the scope of the speculates it would be anticipated utilizing Deep Neural Network [4].

CNN is the most agent model of profound learning [12]. An ordinary CNN engineering, which is alluded to as VGG16, can be found in Fig. 1. Each layer of CNN is known as a component map. A Hidden Markov Model is a finite set of states. Transitions among these states are governed by a set of probabilities called transition probabilities [16]. The component guide of the information layer is a 3D grid of pixel powers for various shading channels (for example RGB). The component guide of any interior layer is an initiated multi-channel picture, whose 'pixel' can be seen as a particular element. Each neuron is associated with a little bit of neighboring neurons from the past layer (responsive field) Different sorts of changes [13] can be coordinated on feature maps, for instance, filtering and pooling. Isolating (convolution) movement tangles a channel grid (learned burdens) with the potential gains of an open field of neurons and takes a nonlinear limit, (for example, sigmoid [14], ReLU) to get last reactions. Pooling activity, for example, max pooling, normal pooling, L2-pooling and neighborhood contrast standardization [15], rundowns the reactions of a responsive field into one worth to create more hearty component portrayals.

Proposed Method

Objectives

1. To help blind people see, not literally but make life a little bit easier for them: This device will help blind people to walk on the road and do their daily works more easily.
2. To collect the dataset of the Images: Collection of different types of dataset of different object required for analysis and classification of the images. To increase the accuracy regarding object recognition a large amount of dataset is required.
3. To capture Image using camera: The important task is to capture the image using camera so that the recognition of the object must be done. Camera may be present on the shirt or corner of the shoulder of the user.
4. To extract features from the Image: Feature Extraction is one of the most important tasks in the object recognition and detection.
5. To break down the components and examples of the Image: Based on the division of picture various elements and examples can be investigated.
6. To recognize the Image.

System Design

The proposal presents object detection and identification device for visually impaired people. In this project, we have proposed a device for object identification with voice feedback for the visually challenged. The proposed fully integrated system has a camera as an input device to feed the real time images of object surrounded by blind people and object detection and identification is done by Mobile-Net SSD algorithm. A methodology is implemented to the recognition of different objects and then provides voice feedback. As part of the software development, the Open CV (Open-source Computer Vision) libraries are utilized to capture image of object. Using Mobile-Net SSD algorithm object is detected from captured images. The proposed system is trained with Pascal and COCO datasets to identify objects from images. Once, the object is recognized, it is made available as an audio output.

Figure 2 shows the system architecture of proposed system. User first has to power on the device. The detected image is identified with Pascal and COCO datasets. Finally, the identified object is sent as an audio output to the user.

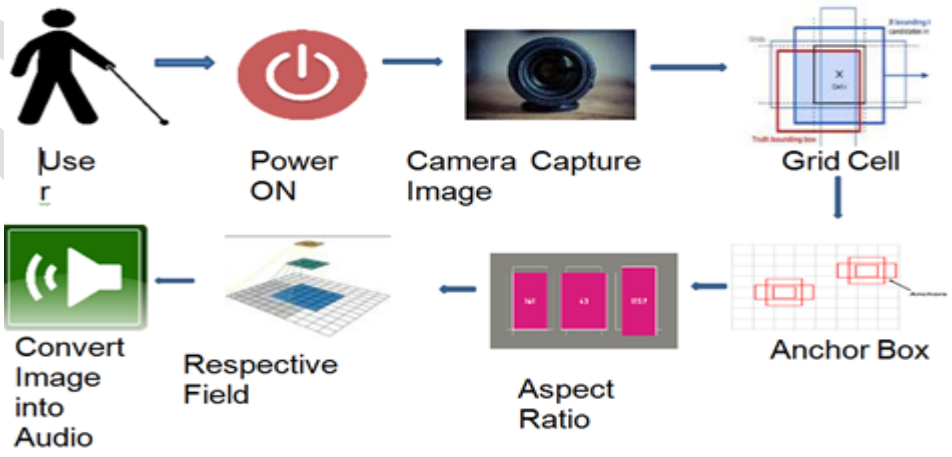

Figure 2: System Architecture

Expected Outcomes

1. The project aims to develop a device which will be able to detect the Objects in the surrounding: Using raspberry pi b+ model and camera module the system will be able to detect objects in the surrounding.
2. Recognition of the detected Object: The object detected by the camera module is then processed by using MobileNetSSD algorithm. The formation of different anchor boxes is done. The probability of the occurrence of the object in that box is calculated. If the probability is more than 20% then that object label is displayed on the screen.
3. Conversion of text output into audio form: Using text-to-speech libraries such as gtts, pyttsx3, engine io the output in the text form is converted into speech. With the help of wireless Bluetooth headphones/earphones/speaker it will be heard by the User (blind person).

Conclusion and Future Work

This Paper present the system for blind people based totally upon object detection and identification. This system makes use of SSD (Single Shot Detection) to identify objects. Object's detection is used to find objects in the real world from an image of the world. The raspberry-pi camera is used to capture real time images from surrounding. To provide voice feedback, the identified image is converted into audio format using GTTS (Google text to speech) module. So, the proposed system will assist the blind people to discover different objects from surrounding and could deliver sound as output to the user. This system is used in actual time object detection. The navigation system is highly priced which isn't always affordable to blind people. So, this project goal is to help blind peoples. Future scope can be, identifying user's known people when they passed by them. Also, to compute the gap between the blind person and every object, on the way to effortlessly recognize how lengthy item from them. To make lifestyles extra simpler of blind human beings the night vision mode will be available in inbuilt camera.

References

1. Z. Zhao, P. Zheng, S. Xu and X. Wu, **Object Detection With Deep Learning**, 30, (2019) pp. 3212-3232 (accessed Jan 30, 2021).Google Scholar
2. Xiao, Y., Tian, Z., Yu, J. et al. **A review of object detection based on deep learning. Multimed Tools** 79, pp 23729–23791 (2020) (accessed Jan 30, 2021)..https://doi.org/10.1007/s11042-020-08976-6
3. https://towardsdatascience.com/12-papers-you-should-read-to-understand-object-detection-in-the-deep-learning-era-3390d4a28891
4. Patil V., Ingle D.R. **An association between fingerprint patterns with blood group and lifestyle based diseases: a review**. 54, (2021), 1803–1839. https://doi.org/10.1007/s10462-020-09891-w
5. Taigman Y, Yang M, Ranzato MA. **Closing the gap to human-level performance in face verification (**2014) , pp 1701–1708.CrossRefView Record in ScopusGoogle Scholar
6. HuangW, Qiao Y, Tang X **Robust scene text detection with convolution neural network inducedmser trees.** (2014) pp 497–511,Google Scholar
7. Zhang L, Lin L, Liang X, He K ,**Is faster rcnn doing well for pedestrian detection? In: European conference on computer vision (ECCV),** (2016**) ,** pp 443–457.Google Scholar
8. Kleban J, Xie X, MaW-Y **Spatial pyramid mining for logo detection in natural scenes. In: IEEE international conference on multimedia and expo (ICME),** (2008) , pp 1077–1080 .Google Scholar

9. A. Dundar, J. Jin, B. Martini, and E. Culurciello, **Embedded streaming deep neural networks accelerator with applications,** IEEE Trans. Neural Netw. & Learning Syst., vol. 28, no. 7, pp. 1572–1583, (2017).CrossRefView Record in ScopusGoogle Scholar
10. A. Stuhlsatz, J. Lippel, and T. Zielke, **Feature extraction with deep neural networks by a generalized discriminant analysis**. IEEE Trans. Neural Netw. & Learning Syst. 23, (2012). pp. 596–608, Google Scholar
11. S. Ren, K. He, R. Girshick, and J. Sun, **Faster r-cnn: Towards realtime object detection with region proposal networks,** (2015), pp. 91–99. Google Scholar
12. Y. LeCun, Y. Bengio, and G. Hinton, **Deep learning**,Nature, 521, (2015). pp. 436–444, Google Scholar
13. M. Oquabetl, **Learning and transferring mid-level image representation using convolutional neural networks,** in CVPR, (2014). (accessed March 30, 2021) .CrossRefView Record in ScopusGoogle Scholar
14. Tilak, G., & Bhaumik, A. A Review on Security and Usability of Graphical User Interface Design.
15. Wadley etl, **Probit analysis: a statistical treatment of the sigmoid response curve,** Annals of the Entomological Soc. of America, 67, (1947),pp. 549–553.Google Scholar
16. K. Kavukcuoglu, R. Fergus, Y. LeCun et al., **Learning invariant features through topographic filter maps,**" , (2009). (accessed Jan 31, 2020) .CrossRefView Record in ScopusGoogle Scholar
17. Swati Warghade, Shubhada Desai, Vijay Patil, **Credit Card Fraud Detection from Imbalanced Dataset Using Machine Learning Algorithm** ,68, (2020), (accessed March 30, 2021). CrossRefView Record in ScopusGoogle Scholar