Journal of Engineering and Applied Sciences 12 (Special Issue 9): 8661-8664, 2017
ISSN: 1816-949X

The Sharp Eye: A Vision for the People Who Lost this Gift

Mrinal Mathur, Pranshu Dixit and S. Chethan
Department of Information and Communication Technology,
Manipal Institute of Technology, Manipal, India

Abstract: Humanity never ceases to amaze us with their advent in technology. Simplifying lives has always been the goal of these new inventions. This manifest effort has worked solely for one purpose that is to "turn the fantasy island into reality". Give the power of technology they deserve. Knowing our surroundings has been metaphorically said recurrently but we are developing an application which gives a vision to the visually impaired and blind. It'll tell them what's around them. The more the wide angle of the camera, more scenes will be captured by the camera. We try to achieve this using 'One Shot Learning'. Instead of loading thousands of training examples, we feed the network a handful of images for it to learn. It makes use of the fact that a 6 years old kid can categorize thousands of objects based on prior knowledge of their respective categories. This application will talk to you on regular intervals, announcing presence of object in your nearby proximity.

Key words: Artificial intelligence, machine learning, neural networks, object recognition, CNNdroid, deep learning, Android

INTRODUCTION

Many state of the art algorithms exist for object detection, face detection, handwriting detection, action detection. Combined them all into a single mobile application is what takes these detections to the next level. And implementing it in a real-time system defines a new dimension. We have planned to use one short learning (Fei, 2003). Imitating the "Human eye" into an application is our motto. We don't want to waste our time in processing, as accidents don't. This application will define a new dimension of world for the visually impaired and blind in their compass (Redmon *et al*., 2016). We will try to implement the existing state of the art algorithms onto our application and give these people a brand new digital vision. Using one shot learning we will try to achieve objects classification. Unlike computer vision where learning requires thousands of parameters and training examples. They fail in real time object classification as going through the large dataset or training example utilizes time, contradicting the real-time definition. In learning a new category of objects, prior knowledge of categories help specially the "variability" in the parameters helps us to stark a contrasting difference and learn better and faster. The application can capture a panoramic shot which is greater than field of view of a human eye which is about 160° by 75°, classify objects and explain the scenario around.

Our aim is to achieve a 360° view and help the disabled imagine a world of their own. Starting with list of objects around and telling the separated distance between them using basic optics will make this application a user friendly and "warn" of obstacles ahead. This application uses face recognition who's basic agenda is to detect face giving us the information regarding sentiments as well, Object recognition which tells us about the object just in front of us given us the accurate recognition to about 0.97 probability given that the image is not take in the dark and not blurry, action recognition which uses the concept of frame traversing because of which it will be able to detect the action in the real time that is happening and then lastly it uses the concept of handwriting recognition which uses 2D convolution (Iandola *et al*., 2013) by converting the image into its corresponding autonomous image and extracting the image out of handwriting recognition will be using neural networks too because in real time we everyone's handwriting might be different.

Literature review: For many of us, face recognition is the basis of recognizing and identifying a person's identity from an image or from a video source. Parkhi *et al.* in their study proposed that the techniques like convolutional neural networks and large training set is being used to make it possible to identify or distinguish any feature in real time. As there are a lot of large datasets available this study deals with more efficient way of

Corresponding Author: Mrinal Mathur, Department of Information and Communication Technology,
Manipal Institute of Technology, Manipal, India

working with the dataset and also by using multiple convolutional neural networks and Bayesian learning by using different network architectures and very deep networks. Object recognition system has been a major challenge as there are a million of objects around the world and making it recognize every object with maximum accuracy is another challenge to look upon. Jain *et al.* (1995) talk about real time object recognition by Meenu (2017) which uses artificial neural networks and back propagation neural networks along with fuzzy logic. This fuzzy logic (Isotani *et al*., 1995) is also an important concept as it allows the system to have many options while recognizing and then select the most accurate and classify into its respective categories. Redmon *et al.* (2016), unlike r-CNN train a CNN to predict regions of interest instead of using selective search which often gives less accuracy. Handwriting recognition is not a new field that has been explored but there are various methods on how to work upon this problem. Back propagation networks to hand-written recognition, Cun in study reduces the complex pre-processing of detailed engineering of images, rather it consist of black and white pixels and the characters are also well separated from the background of study or any screen. The recognition part remains entirely on the performance of multi-network layers (Kivela *et al*., 2014) because of which the network could be trained in the low level representation.

MATERIALS AND METHODS

The sharp eye will be using the concept of convolutional neural networks and artificial intelligence to identify items. This will all be done in real time so any moment it can recognize items. Using convolutional networks, these can be used in computer vision to recognize items that may not always appear the same way, like animals, people and or common objects like flowers, etc. We will be implementing CNN in android in using CNN droid library (Oskouei *et al*., 2016) for execution of trained convolutional neural network on any android phones.

The main highlights of CNN droid that we will be using will be support for nearly all the CNN layer types and compatible with CNN modes trained by the common library which are Cafe and Torch.

- Create a renderscript object (render script provides a platform-independent computation engine that operates at the native level. Use it to accelerate your apps that require extensive computational horsepower (Iandola *et al*., 2013)
- Construct a CNN droid object and provide NetFile location address
- Prepare your input to the network
- The input can be single or batch of images
- Call the compute function of the CNN droid library and get the result of the CNN execution as an object when the computation is finished

CNN makes the learning rate better by the scale of +23 with much more accuracy rather than outdated modeling techniques like Naive Bayes and decision trees.

Since, it needs training as well as a testing set to learn more efficiently and it might happen that the object or action is not recognized by this application and then we will be giving the user an option that allows the user to take a picture of that object and appends it to already given list of items.

We will be using trace transform method and Elastic Bunch Graph Matching (EBGM) method for proper recognition which will be implanted using machine vision (Field, 2015) this attempts in integrating the existing technologies in a very sophisticated way that can we use for industrial purposes as well. In this we can go step by step. Firstly, we will use machine vision by clicking an image by using cameras and lenses. There are various software's for digital image processing as well which is used to extract the relevant information based on classification and making decision according to various categories and classes. For this various method will be used such as.

Equipment: Automatic inspection system includes lighting of image, a camera, processor, software and output devices.

Imaging: This will be used to separate either by main image processing or by using a smart camera. When this is separated, the connection may be used to specialized intermediate hardware and a frame grabber.

Image processing: After acquiring the image, it is processed. MV uses multiple stages of processing which includes stitching/registration, filtering, thresholding, pixel counting, segmentation, edge detection, pattern recognition, neural net, etc.

Artificial intelligence will be used to understand our camera's input and find out what it sees. Soon, it can also be used to understand complex scenes, the position of image within it, so, it can describe the relationships between items it sees such as "dog near a lamp post". This is implemented properly using machine vision as

well. This is done by dividing the original picture into sub-parts and further into different points of interest and matches each of them to the object or entity. This should be particularly divided into 8×8 block which allows us to see the best result and hence, the screen orientation of the image will be compromised. The algorithm of an image divided into blocks is given:

- Generate a test image
- Define the window size
- Crop out the window and calculate the histogram

RESULTS AND DISCUSSION

After getting the image as an output will then be matching it to the testing set which will then give the result accurately. This all will be implemented in only android application and so, the framework that we used to implement this is called deep learning 4j which is used by little effort. Deep learning 4j is a method of allowing us to access deep neural networks from shallow or small networks which are called layers or many of times also called 'hidden layers'. This combines a lot of Bayesian systems, convolutional networks and recurrent convolutional networks which can also be looked forward by using frameworks by using various big data techniques like Hadoop, etc. We will be using Torch7 of android which provides a MATLAB-like enviroment, for machine learning algorithm.

```
CaffeMobile::~CaffeMobile()
{net_.reset(); }
        voidCaffeMobile::SetMean(const vector<float>&mean_values) {
         CHECK_EQ(mean_values.size(), num_channels_)
          <<"Number of mean values doesn't match channels of input layer"
         cv::Scalar channel_mean(0)
         double *ptr = &channel_mean[0]
         for (inti = 0; i<num_channels_;++i) {
         ptr[i] = mean_values[i]
         }
`
            mean_ = cv::Mat(input_geometry_, (num_channels_ == 3 ?
         CV_32FC3 : CV_32FC1)
         channel_mean)
         }
```

This will also be using LSTM. The latter contain information outside the normal flow of the recurrent network in a gated cell (Bell *et al*., 2016). Like a computer's memory, a cell can store information. Information can also be written to, or read from a cell. Decisions regarding what info to store, read time are made by the cell itself. They also write and erasure, via. gates that open and close. These gates are however, analog, implemented with element-wise multiplication by sigmoid which are all in the range of 0-1. Analog has the advantage over digital as it has better mathematical properties such as differentiability and hence is a better option for back propagation. In this, since, we will be able to give the user a chance to upload or train those objects that have not been trained. All the images can be stored in a cloud storage which will send an HTTP request with our image and we will receive the description of its content. This way, it can be circulated all over the world. In this there will be some objects that might be offending for others which we will be by default mark it as offensive and no other user will be able to train it in any way.

CONCLUSION

The sharp eye is a very simple application. It takes pictures and the system will search for this using CNN. This application will surely have some limitations like it may not always correctly distinguish between genders for instance and while it can recognize certain very strong and distinct facial expressions such as a person smiles broadly or someone looking very angry it's not yet very sensitive. Application's deep learning engine will make it a better identifier than other API's available in market. We hope that to be the main asset of our business and to use that algorithm that not only recognizes objects that can be products but objects that should do with navigation, to help not only the visually impaired but everyone from humans to machines in navigating and recognizing the world. That's the core mission of the sharp eye.

REFERENCES

Bell, S., C.L. Zitnick, K. Bala and R. Girshick, 2016. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, IEEE, Las Vegas, Nevada, USA., ISBN: 978-1-4673-8852-8, pp: 2874-2883.

Fei, L.F., 2003. A bayesian approach to unsupervised one-shot learning of object categories. Proceedings of the 9th IEEE International Conference on Computer Vision, October 13-16, 2003, IEEE, Nice, France, ISBN:0-7695-1950-4, pp: 1134-1141.

Field, M.S., 2015. Machine vision and sensing with an Android. Master Thesis, University of Southern Queensland, Toowoomba, Queensland.

Iandola, F.N., D. Sheffield, M.J. Anderson, P.M. Phothilimthana and K. Keutzer, 2013. Communication-minimizing 2D convolution in GPU registers. Proceedings of the 20th IEEE International Conference on Image Processing (ICIP), September 15-18, 2013, IEEE, Melbourne, Victoria, Australia, ISBN:978-1-4799-2341-0, pp: 2116-2120.

Isotani, S., A.R.D. Albuquerque, H. Sidaoui, M. Muratore and O.B. Martins, 1995. A 1-D pattern recognition system based on fuzzy logic. Proceedings of the IEEE 21st International Conference on Industrial Electronics, Control and Instrumentation Vol. 2, November 6-10, 1995, IEEE, Orlando, Florida, USA., ISBN:0-7803-3026-9, pp: 1058-1061.

Jain, R., R. Kasturi and B.G. Schunck, 1995. Machine Vision. McGraw-Hill, New York, USA., ISBN:9780070320185, Pages: 549.

Kivela, M., A. Arenas, M. Barthelemy, J.P. Gleeson and Y. Moreno et al., 2014. Multilayer networks. J. Complex Netw., 2: 203-271.

Meenu, P.J., 2017. Real time object recognition system. Intl. J. Theor. Appl. Mech., 12: 241-266.

Oskouei, S.S.L., H. Golestani, M. Hashemi and S. Ghiasi, 2016. Cnndroid: Gpu-accelerated execution of trained deep convolutional neural networks on android. Proceedings of the ACM Conference on Multimedia, October 15-19, 2016, ACM, Amsterdam, Netherlands, ISBN:978-1-4503-3603-1, pp: 1201-1205.

Redmon, J., S. Divvala, R. Girshick and A. Farhadi, 2016. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, IEEE, Las Vegas, Nevada, USA., ISBN: 978-1-4673-8852-8, pp: 779-788.