Augmented Reality Artificial Intelligence American Sign Language/ Visually Impaired Classification Smart Glasses

Hashim Farooq

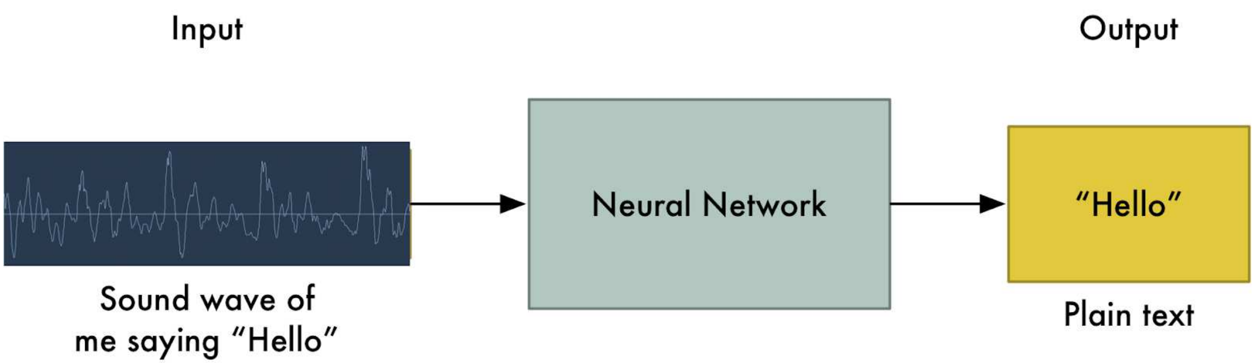
Introduction:

Communication is vital part of day to day life that we all take for granted. Being able to talk to anyone we choose is a privilege that some don't have. Hearing impaired have a lack of people they can communicate with. Most of them can only "talk" to people that "speak" ASL or other forms of gesture languages. Everyone should have the right to speak to whomever they choose and language should not be a limitation. There are over 1 million people in America that use ASL (NFB.org) and all these people have missed the opportunity to meet new people, make new connections, make new relationships all because of communication or lack of. I did some research to see the solutions currently on the market and all of the current solutions are expensive and lack a lot important features that would help with natural conversations between the hearing impaired and non-hearing impaired.

This motivated me to help make a device that was able to give a better quality of life to people with these disabilities. I wanted to make something to also help out the visually impaired.

I came up with the idea of building smart glasses as they could be convenient and offer a sort of natural conversation. The glasses will be able to translate speech into text which will be displayed in front of the glasses of the hearing impaired and they will classify the ASL into text which will be displayed in front of the non-hearing impaired. For the visually impaired the glasses will use AI image classification to detect objects around and tell the user of its surroundings.

The goal of this project is to build a device that is inexpensive and effective to help bring a better quality of life.



Process:

This project needs 3 programs: the voice to text classifier, the ASL to text classifier, and the object classifier. All the programs were coded in Python as I have pervious knowledge of the language and experience working with it.

To make the voice to text classifier I wanted to it simple and cost efficient, so I used Google voice recognition and wrote simple commands in python to show the text and let it be displayed on the screen. The main purpose of using google was that it already has a well-developed library for voice recognition and trying to remake one would be very expensive and could take years to get results.

Making the ASL to text classifier was harder than I had anticipated mostly because this was my first machine learning and neural network project. This project was broken down into 3 files, file 1 was the collect data file, file 2 was the training file and file 3 was the main display file. In the first file would collect the data and store it into a local folder, the second file takes the data and trains the model to help it better understand the ASL and the third file predicts the hand gestures. The model has currently been trained to detect number gestures but soon will be trained with other gestures, due to lack of time and lack of computer memory I was unable to further train the data files. In order to reduce data usage, the collect data file converts the collected data into a 15x15 pixel image and inverts the colors. This does reduce the data and helps increase the speed of the training process, but it reduces the accuracy of the predictions. The training file has over 1000 images of each gesture taken by me. The more images in the data files the faster and more accurate the model becomes.

The same process was used for the object detection program. Collecting the data for this program took a long time as taking 1000 photos of various kinds of objects was difficult for this program I took 100 photo myself in different environments and different lighting conditions to assure the best data was collected, for the 900 rest I sourced the data on Google Images. The training of the model took a very long time just over 1 week of training data.

In order to run these programs on smart glasses, I used a Raspberry Pi Zero as it was cost efficient, compact and most importantly powerful. I then uploaded the files onto the Pi and made a main interface.

The glasses are to be build using a 3D printer, but due to the lack of a 3D printer I was unable to construct the glasses.

Results:

I could not construct the glasses due to lack of a 3D printer, so I mounted all the parts on a pair of sunglasses to get some sort of results. The glasses performed almost perfect in day light and in well-lit locations but when it came to low lighting the glasses almost did not work as well. There is a simple solution to this problem and that is to use a night sight camera. Using this camera would enhance the ability of the glasses and does not really affect the cost either.

The glasses use the camera to analyze the area. When someone appears in front of the glasses the glasses will detect the person and convert their speech or ASL into text which will then appear in front of the user.

On the second application the glasses classify the objects into speech and vibrations. The user looks around and the camera of the glasses detects the objects closest to the camera. If the user id crossing a street and there is an object in front of them (3-8feet) then the glasses will trigger a sequence of vibrations.

Conclusion:

I spent a day walking around in these glasses and found that they were convenient in other ways too not only as a way of breaking the communication barrier but also as a way navigating and viewing the whole world in a different perspective. The AI was able to detect random objects that I was preciously unaware of. The total cost of the glasses were just $200 (cad) but if the materials were bought in wholesale the piece could've been dramatically lower.

With more training and programming these glasses will be able to help the hearing and visually impaired live a better quality of life.

There are endless future applications for smart glasses like these. One I'm very curious and wanting to try is to make a program for the glasses for people who have Alzheimer's and dementia. The glasses would have data stored of the patient's personal life and display it every morning, so he/she gets a recap of their life and who they are. The glasses will also be able to identify the persons family members and friends. It will use CNN to detect the faces and if the face is stored in the data file it will be able to identify the person and display a short summery of the person and how they are related. The glasses will be able to recive new data by the family members updating their profiles and new events so that the patient is aware of the events. These glasses will be able to provide a better quality of life of these patients and make it easier to live their day to day lives. It's tough to live life not knowing how you lived it.

References:

1. ***Ming Liang, Xiaolin Hu***; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3367-3375
2. Y. Yao, Y. Wei, F. Gao and G. Yu, "Anomaly Intrusion Detection Approach Using Hybrid MLP/CNN Neural Network," Sixth International Conference on Intelligent Systems Design and Applications, Jinan, 2006, pp. 1095-1102.
3. F. Chen and M. R. Jahanshahi, "NB-CNN: Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion," in IEEE Transactions on Industrial Electronics, vol. 65, no. 5, pp. 4392-4400, May 2018.
4. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," pp. 211--252, 2015.
5. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in NIPS, 2012, pp. 1097--1105.
6. M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in ECCV, 2014, pp. 818--833.
7. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," arXiv preprint arXiv:1409.4842, 2014.
8. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," arXiv preprint arXiv:1512.03385, 2015.
9. C. Zhang, P. Li, G. Sun, Y. Guan, B. Xiao, and J. Cong, "Optimizing fpga-based accelerator design for deep convolutional neural networks," in Proceedings of ISFPGA. ACM, 2015, pp. 161--170.
10. T. Chen, Z. Du, N. Sun, J. Wang, C. Wu, Y. Chen, and O. Temam, "Diannao: A small-footprint high-throughput accelerator for ubiquitous machine-learning," in ASPLOS, vol. 49, no. 4. ACM, 2014, pp. 269--284.
11. Y. Chen, T. Luo, S. Liu, S. Zhang, L. He, J. Wang, L. Li, T. Chen, Z. Xu, N. Sun et al., "Dadiannao: A machine-learning supercomputer," in MICRO. IEEE, 2014, pp. 609--622.
12. D. Liu, T. Chen, S. Liu, J. Zhou, S. Zhou, O. Teman, X. Feng, X. Zhou, and Y. Chen, "Pudiannao: A polyvalent machine learning accelerator," in ASPLOS. ACM, 2015, pp. 369--381.
13. Z. Du, R. Fasthuber, T. Chen, P. Ienne, L. Li, T. Luo, X. Feng, Y. Chen, and O. Temam, "Shidiannao: shifting vision processing closer to the sensor," in ISCA. ACM, 2015, pp. 92--104.
14. ***Yikang Li, Wanli Ouyang, Xiaogang Wang, Xiao'ou Tang***; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1347-1356
15. L. Deng, J. Li, J.-T. Huang, K. Yao, D. Yu, F. Seide, M. Seltzer, G. Zweig, X. He, J. Williams, Y. Gong, , and A. Acero, “Recent advances in deep learning for speech research at Microsoft,” in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, May 2013.
16. T. N. Sainath, B. Kingsbury, B. Ramabhadran, P. Fousek, P. Novak, and A. Mohamed, “Making deep belief networks effective for large vocabulary continuous speech recognition,” in ASRU, 2011.
17. D. B. Sam, S. Surya and R. V. Babu, "Switching Convolutional Neural Network for Crowd Counting," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 4031-4039.

18. L. Cavigelli, P. Hager and L. Benini, "CAS-CNN: A deep convolutional neural network for image compression artifact suppression," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, 2017, pp. 752-759.
19. R. Vinayakumar, K. P. Soman and P. Poornachandran, "Applying convolutional neural network for network intrusion detection," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 1222-1228.
20. T. Matsumoto, L. O. Chua and T. Yokohama, "Image thinning with a cellular neural network," in IEEE Transactions on Circuits and Systems, vol. 37, no. 5, pp. 638-640, May 1990.
21. C. Wu, W. Fan, Y. He, J. Sun and S. Naoi, "Handwritten Character Recognition by Alternately Trained Relaxation Convolutional Neural Network," 2014 14th International Conference on Frontiers in Handwriting Recognition, Heraklion, 2014, pp. 291-296.