Machine Learning based Smart Glasses for Visually Impaired People

Team: **DATA ROOS**

- Vijay Agavinti
  (Vactp@umkc.edu)
  Masters in computer science
  University of Missouri Kansas City

- Chandu Chandra
  (vcypr@umkc.edu)
  Masters in computer science
  University of Missouri Kansas City

Albishri Ahmed
(aa8w2@mail.umkc.edu)
Python and Deep Learning
University of Missouri Kansas City
School of Computing and Engineering

ABSTRACT:

In the world of colors, It's sad that some people can't see because of some kind of visual impairment. So, we decided to build novel smart glasses based on a machine learning model. These glasses can be capable of recognizing the images and produce output in audio format, so that the person with the visual impairment is able to hear and visualize what is in front of them

---

INTRODUCTION:

Machine learning based smart glasses for visually impaired people, Main goal is to detect an object, recognize it and produce the output in audio format. This model uses a pretrained model by google net which has an accuracy of 95 percent.

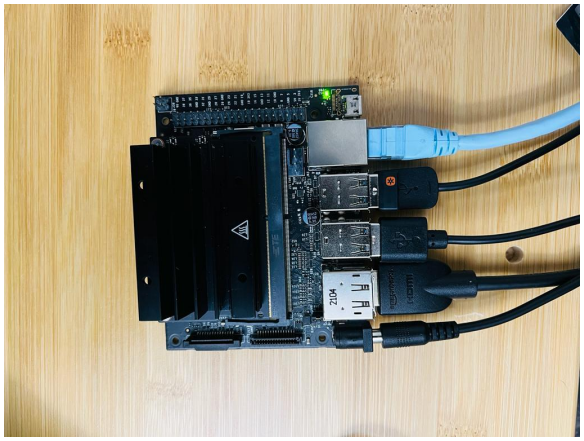
First of all we need to install the OS with which we can operate the nano. After that we have to connect the required components to the nano, such as camera and speakers, Since nano didn't have an inbuilt wifi feature, we have to use either wired ethernet or we can connect wireless wifi module to the nano which has a port under the heatsink. We need a display to visualize while building the model. Unlike laptops it won't have a display, keyboard and trackpad, So keypad and mouse also needed to operate the jetson nano module.

Nano can support upto 5v and 1.4 amp, we have to make sure that the power supply has to be proper, otherwise the module may suffer with thermal throttling.

Hardware:

i) Jetson nano 2GB

Jetson nano is a Developer Kit, it is a compact, powerful computer that allows you to run many neural networks in parallel for image classification, object recognition, segmentation, and speech processing. All of this is contained within a user-friendly platform that consumes as little as 5 watts.



ii) Microsoft web camera
iii) Speakers
iv) Display
v) keyboard and mouse
vi) power adapter

Main Software Used:

i) Googlenet

GoogLeNet was built on the "Inception" deep convolutional neural network architecture, which was responsible for establishing a new standard for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC 2014).

ii) ubuntu :The operating system that is used to operate the Nano module.

iii) gTTS: Google text to speech, This module allows us to convert the text into audio format

Step 1: Installing OS

We need to install the Operating system from the NVIDIA Jetson nano website. After installing the software in the Micro SD card we have to insert the SD card into the Nano.

Step2: Connections

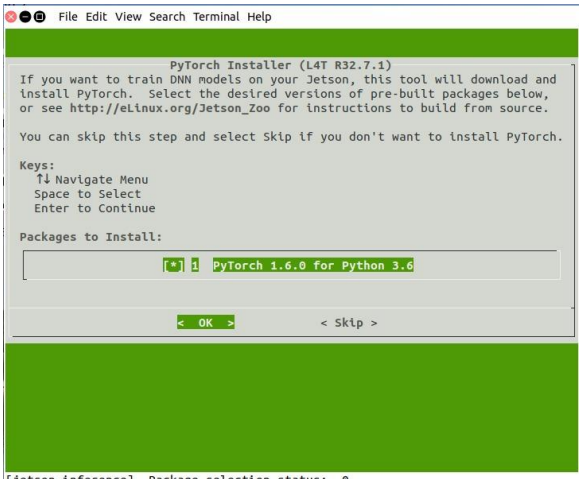
Step 3: Installing compiler and required libraries: There are few libraries and applications to be installed in order to detect the image and convert that detected output into the audio format.

- Python3

Command: *sudo apt-get install python3*

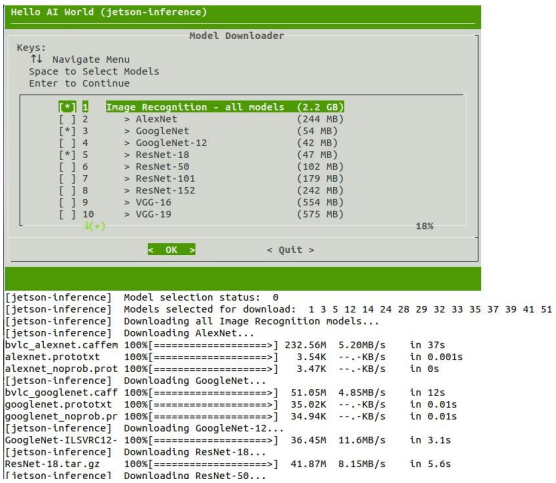
- VScode
- Jetson interface: This consists of all the libraries that are required for the object detection

Command: *sudo apt-get install jetson.interface*



- Jetson utils: This have all the libraries that are required for image recognition

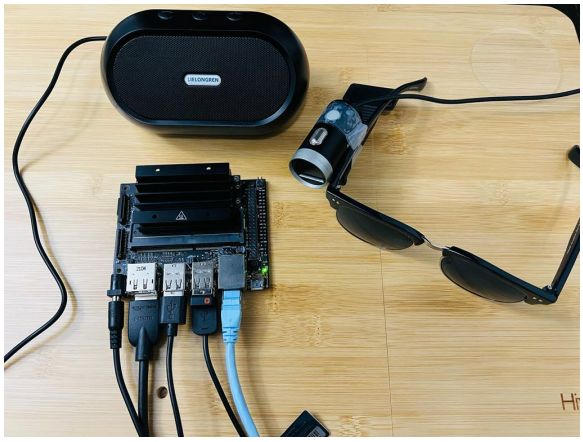
Command: *sudo apt-get install v4l.utils*



-gTTS: Text to speech conversion is one of the smart glasses' most essential features. We installed gTTS in order to complete this assignment (Google Text-to-Speech). It's a Python module that works with the Google Translate API [13]. Many features of gTTS include the ability to convert any length of text to speech, enable error pronunciation using

customisable text preprocessors, and support multiple languages with the ability to retrieve them when needed.
Command: *sudo apt-get install gTTS*

Step4: Model



We are using a pretrained model by google i.e.Google net, it has an accuracy of 95%. This model uses deep convolutional neural networks. Which have 18 layers which includes convolution layers, max pooling layers, inception layers, linear layer and dropout layer. This model uses Softmax activation function. Parameters and layers order has been shown in fig1.

Step5:

Using the web camera, Smart glasses capture an image and recognize it and produce output through speakers.

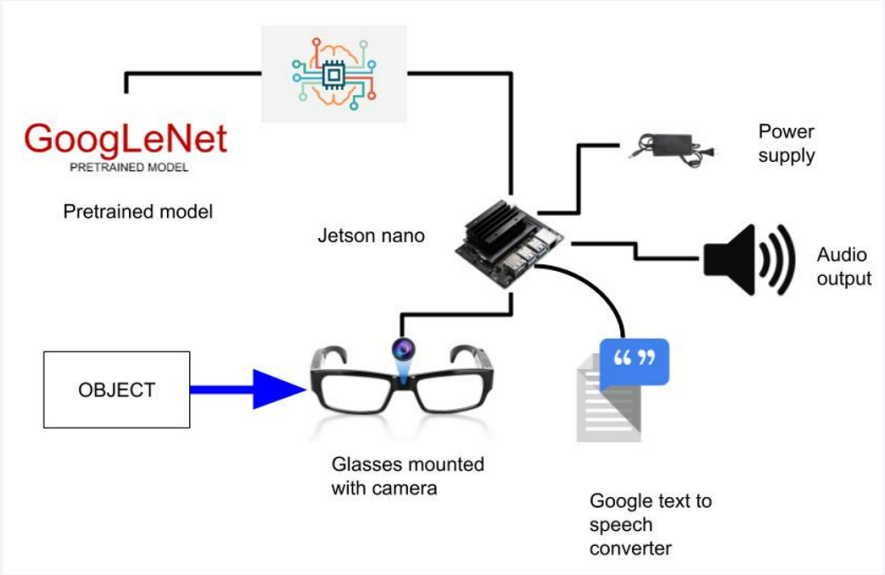
Team	Year	Place	Error (top-5)	Uses external data
SuperVision	2012	1st	16.4%	no
SuperVision	2012	1st	15.3%	Imagenet 22k
Clarifai	2013	1st	11.7%	no
Clarifai	2013	1st	11.2%	Imagenet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no

Fig: Comparison of other models with Google net

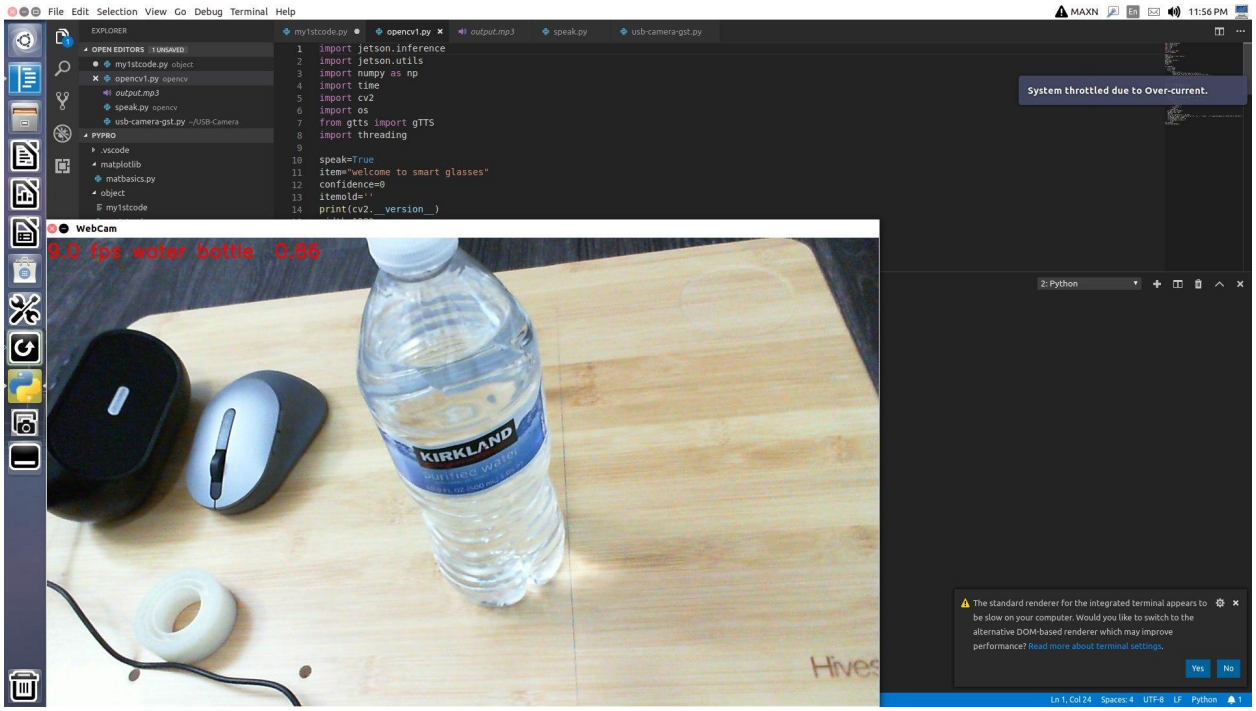
type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Fig 1: This figure represents the model summary of the google net. Which helped in object detection and recognition. As shown in the figure, Model consists of 18 layers and softmax as an activation function

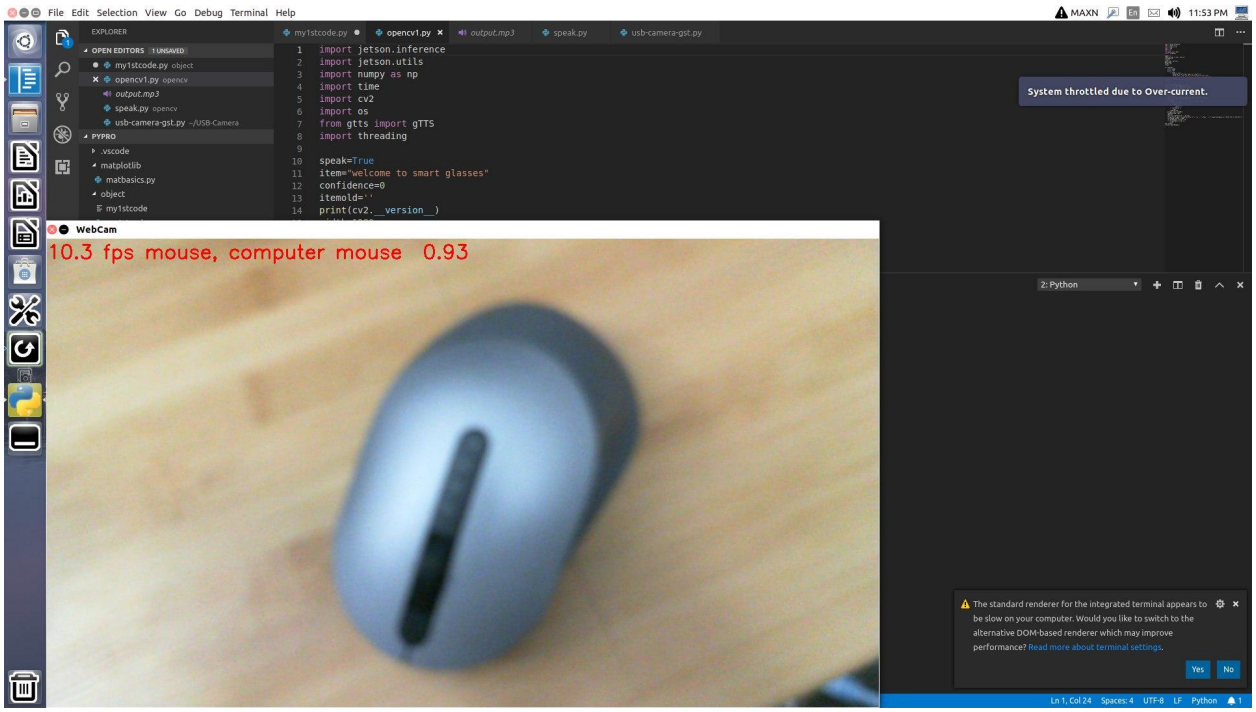
DESIGN:



OUTPUT:



This image represents the output screenshot taken while running the model, In the above image the model predicted that it is an water bottle with 86% of accuracy and in the below image the model predicted that it is an mouse with 93% of accuracy even though the image is not clear.



APPLICATIONS:

This model is designed for visually impaired people, So, The main intention and application for this project is to at least provide the data that can helps in visualizing the things that are in front of the person in the form of audio

Future work:

We planned to include the text recognition feature to the smart glasses which can be used to read the test from the image captured by the camera.

We are also planning to implement the new feature in the future which allows glasses to remember faces and objects from which are not in the dataset, i.e.If smart glasses encounter a new object then the user will be able to add the name of the object to the memory of smart glasses. So that glasses are able to detect an object if we ask it again.

Conclusion:

Since these smart glasses use a pre-trained model, it has a high accuracy of 95%, Which surpasses the previous model with an accuracy of 93%. This model will be further developed as we mentioned in the future work.

REFERENCES:

[1] H. AlSaid, L. AlKhatib, A. AlOraidh, S. AlHaidar and A. Bashar, "Deep Learning Assisted Smart Glasses as Educational Aid for Visually Challenged Students," 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), 2019, pp. 1-6, doi: 10.1109/ICTCS.2019.8923044.

[2] W. -J. Chang et al., "A Deep Learning Based Wearable Medicines Recognition System for Visually Impaired People," 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2019, pp. 207-208, doi: 10.1109/AICAS.2019.8771559.

[3] W. -J. Chang et al., "A Deep Learning Based Wearable Medicines Recognition System for Visually Impaired People," 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2019, pp. 207-208, doi: 10.1109/AICAS.2019.8771559.

[4] 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), 2022, pp. 621-626, doi: 10.1109/MECON53876.2022.9751975.