**Proposal**

1. **Domain background:**

Machine learning models has been the flavor of artificial intelligence field especially convolution neural network which exhibit a very prominent result in detecting patterns in images.

As engineers we have a great responsibility towards people to help them and making their life easier, for example helping blind people to easily interpret their surrounding and to easily interact with the surrounding environment.

1. **Problem statement:**

Blind people have many difficulties to identify objects and they need to be assisted through daily life and navigating through homes.

Indoor recognition is a challenging problem as most recognition scene models were specially made for outdoor recognition and have poor performance in indoor classification.

The aim of this project is to help blind people easily navigate through homes by classifying the type of each room in a house whether it’s a bedroom ,kitchen ,bathroom ,dining room and other rooms that may exist in a home.

1. **Datasets and inputs:**

The dataset contain 10 classes [‘bathroom’ , ’bedroom’ , ’children\_room’ , ’closet’ , ’corridor’ , ’dining\_room’ , ’kitchen’ , ’living\_room’ , ’pantry’ , ’staircase’]

Total number of images =2301

* Training set =1472
* Validation set=368
* Testing set=461

**Citation:**

Scene Parsing through ADE20K Dataset. Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso and Antonio Torralba. Computer Vision and Pattern Recognition (CVPR), 2017.

Semantic Understanding of Scenes through ADE20K Dataset. Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso and Antonio Torralba.arXiv:1608.05442.

Quattoni, and A.Torralba. Recognizing Indoor Scenes. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009

1. **Solution statement:**

In my opinion, the best solution for classifying image is to use the a Convolution neural network as it well known for its great performance in detecting complex patterns and shapes in images.

I intend to use a pre-trained model like Resnet50 ,Inception ,VGG16 ,Mobilenet to extract the best features in the frames and then fine tune the model to be suitable for the problem proposed, then do some coding to make classifying prediction audible by text to audio conversion.

There are some difficulty that may arise from using large pre-trained model like resnet50 and Inception as they need a huge computation and time for transferring the input data to be fed to the Softmax dense layer ,so I decided to use a light model like Mobilenet to be able to classify the objects in real time without losing much accuracy.

1. **Benchmark Model:**

The benchmark model will be a random classifier model ,and the solution for the problem will be a based on Resnet50 or Mobilenet models.

1. **Evaluation Metrics:**

The evaluation metric is based on accuracy ,which is being calculated by the given formula

Accuracy=(True positive + True negative)/(True negative + True positive +False positive +False negative) which mean that we are looping through all the test dataset and compare the result of the predicted frame to it’s true value ,if the model predicted right we give a point and if not no point given, in the end we accumulate all the points and divided it by all the tested frames.

Another evaluation that also need to be considered is how fast the program can takes frames and process it ,because real time application need to fast and interactive.

1. **Project Design:**

I am going to discuss the steps that I would attempt to solve the problem

1. Understanding the problem and collecting the suitable data for the problem.
2. Loading the data and categorize each class type
3. Implement the random model classifier and calculate the accuracy of the model
4. Implement a function that takes an image path and rescale the Image and convert it to array ,then expand the dimension of the array to be suitable for the model based input dimension like “Resnet50” or “Mobilenet”
5. Convert all the image using the function in step 4 ,then preprocess the images using Keras “preprocess\_input” function ,then extract bottleneck features using Resnet50 or Mobilenet models that already pre-trained on the Imagenet dataset
6. splitting the data into training, validation and testing
7. Create a new model to be able to classify the proposed problem
8. Train the model on the bottleneck features and calculate the accuracy and the run time of each model
9. Compare the result to decide which model is better and suitable for real time application
10. Make function to read stream of images using my laptop camera and classify the object and read it aloud, to help blind people guide their way in homes