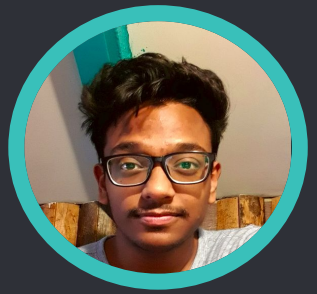
Application of Deep Learning for Indoor Scene Understanding by Robot.

Group Members

Milan Ashvinbhai Bhuva
IIT2018176

Manav Kamlesh Agrawal
IIT2018178



Mohammed Aadil
IIT2018179

Ankit Rauniyar
IIT2018202

Kisalaya Kishore
IIT2018079

Problem Statement and Objective

1

- The objective of this project is to develop a model which can provide assistance in the recognition of Indoor scenes.
- The end goal of this project is to be able to successfully build an application which will provide assistance to the visually impaired while navigating indoors. The model devised by us could be used for multiple purposes which are stated below.

2

Literature Survey

A brief summary of the Base Paper.

Base Paper : Fusion of Transfer Learning Features and Its Application in Image Classification

Year: 2017

Method: Multiple CNN i.e VGG16, Inception-V3, AlexNet, were used to extract features, followed by dimensionality reduction using PCA. The final set of features were used to train the SVM classifier. This was performed on different datasets.

Dataset:

MIT67 - http://web.mit.edu/torralba/www/indoor.html
Sun397 - https://vision.princeton.edu/projects/2010/SUN/
Caltech256 - http://www.vision.caltech.edu/Image_Datasets/Caltech256/
Caltech101 - http://www.vision.caltech.edu/Image_Datasets/Caltech101/
CIFAR10 and CIFAR100 - https://www.cs.toronto.edu/~kriz/cifar.html

Results:

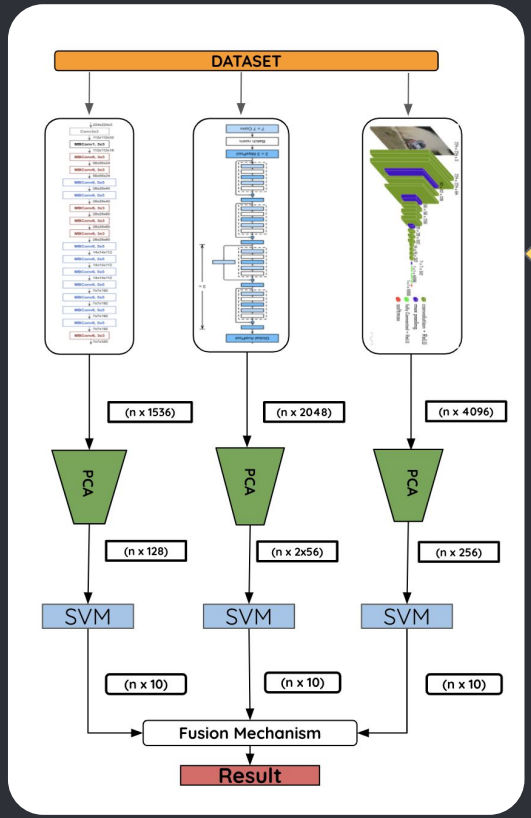
Dataset	ACC(%)
MIT67	75.10
Sun397	61.70
Caltech256	86.30
Caltech101	95.80
CIFAR10	91.70
CIFAR100	75.80

3

Proposed methodology

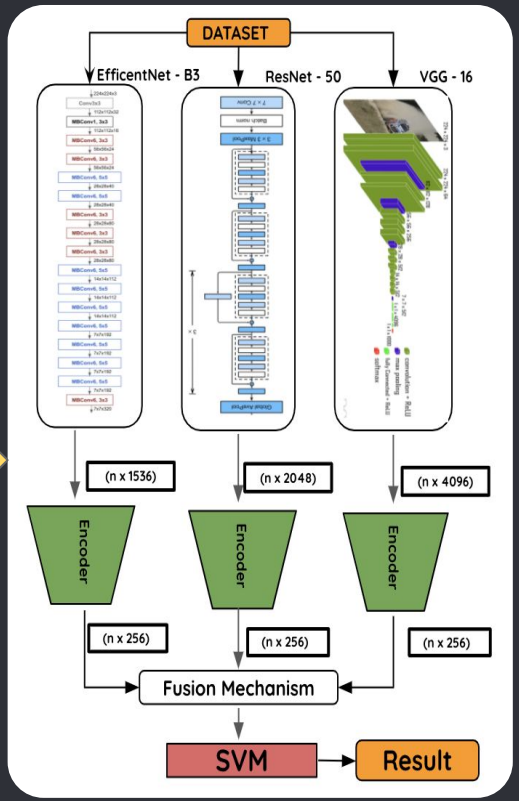
The Inner workings of our model

Visual Representation of Our Methodology



MMFM

HEM



Methodology

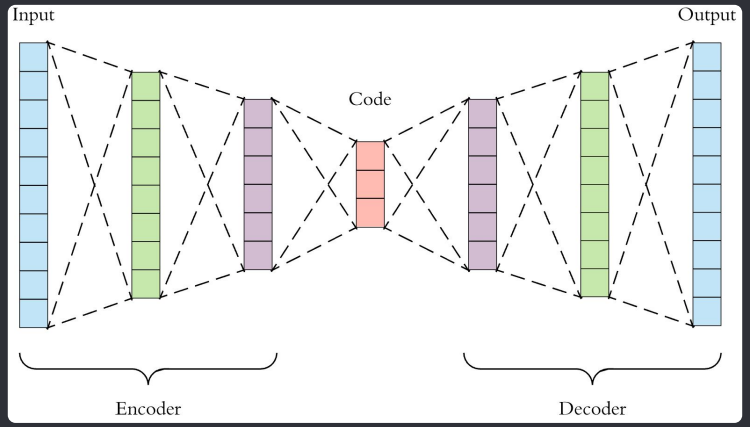
- The major difference between the MMFM and HEM model lies in the algorithm used for dimensionality reduction and order of fusion mechanism and SVM classification.
- The MMFM model uses PCA for dimensionality reduction and performs SVM classification on the features obtained from the neural networks individually, followed by fusion mechanism. Fusion mechanism combines the results obtained from each of the SVM classification to perform a more sophisticated classification.
- The HEM model uses autoencoders for dimensionality reduction and performs feature fusion on the three set of features obtained after dimensionality reduction, i.e. combines the three set of features. Subsequently, SVM classification is performed on the fused set if features.

Contributions

- Proposed a MultiModal leveraged with Fusion Mechanism.
- This model can predict faster than the base paper model.
- Refined hybrid neural model of the base paper by replacing the linear dimensionality reduction with non linear dimensionality reduction technique.
- We implemented a GUI for both image and live feed classification.

Encoder - HEM

- In HEM model , for dimensionality reduction we've used a non-linear dimensionality reduction technique.
- Encoder are useful to attain a certain desirable stage for any given input.



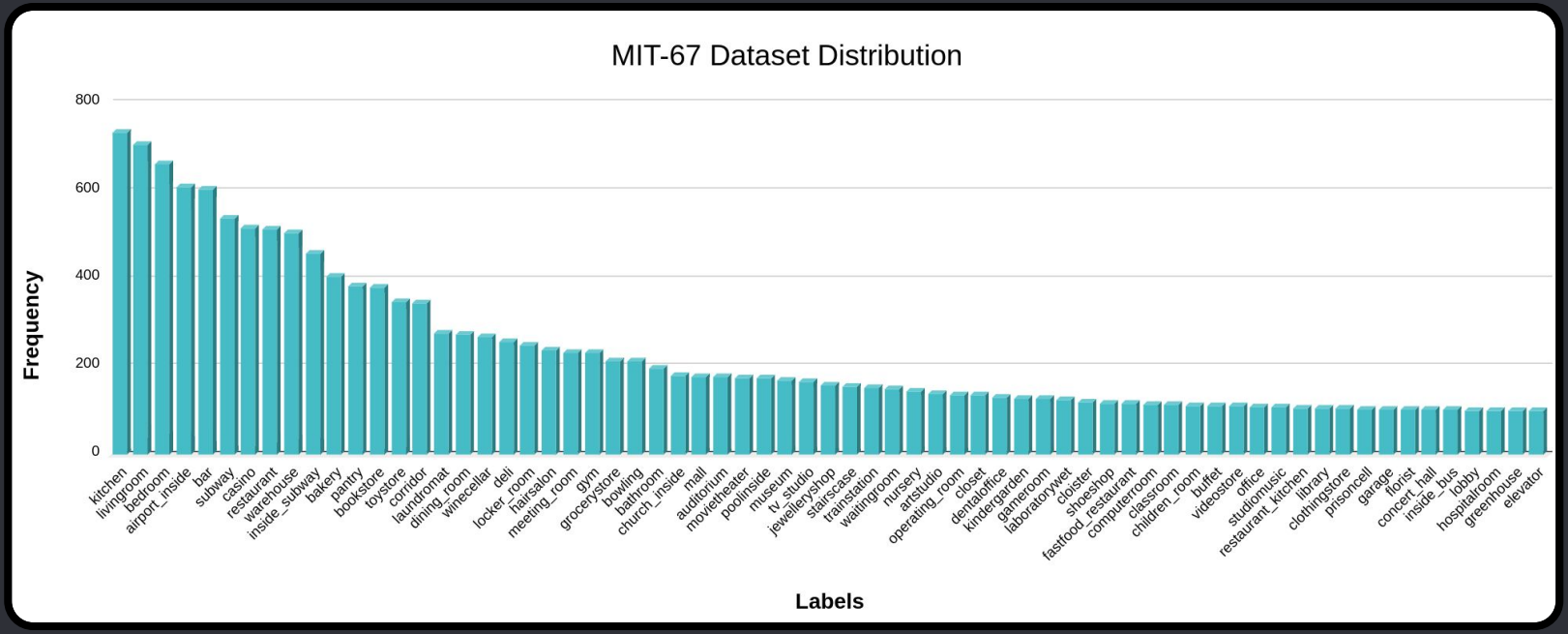
- Basically, we've taken half of the part from auto encoder to achieve our goal shape.
- Encoders are nothing but a NN for which it learns weights and biases for a given input which can be used to regain the original state of the input using decoders.

4

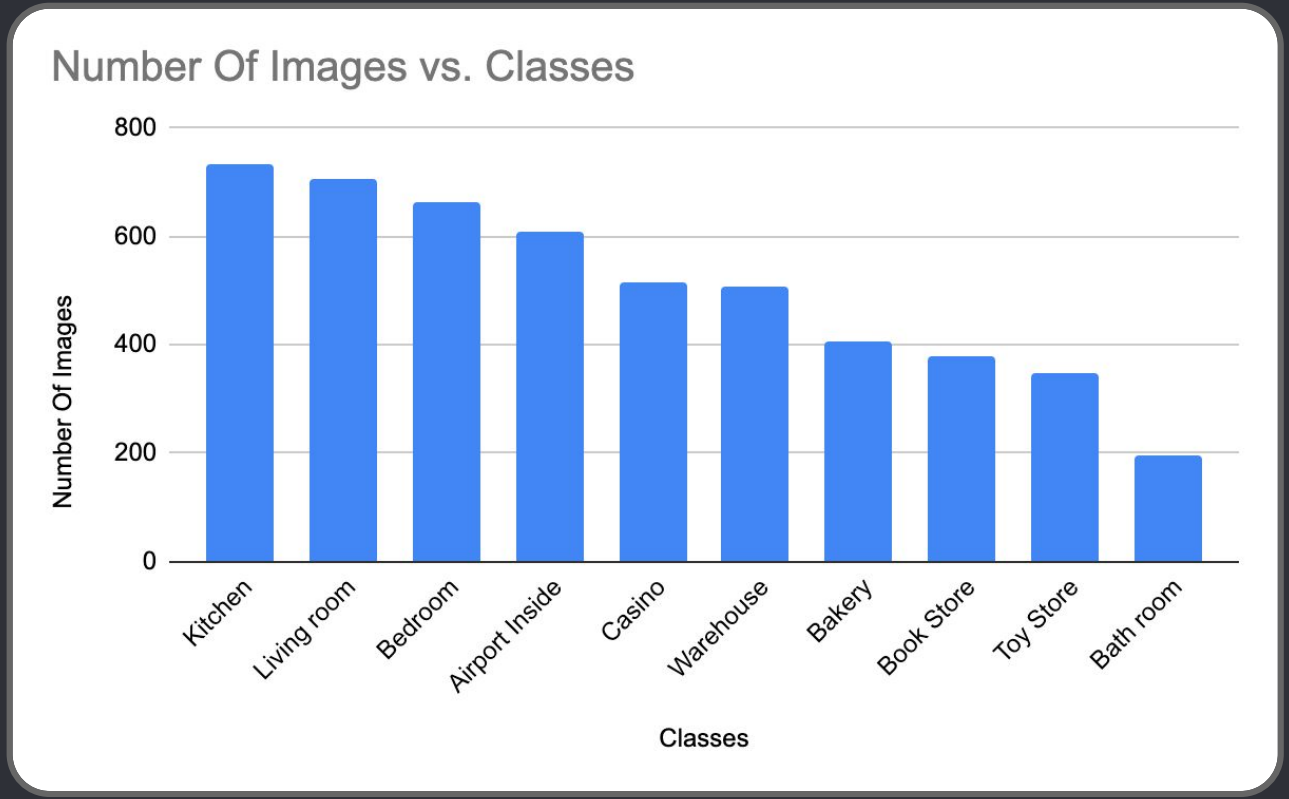
Dataset

The backbone of this project

- We are using the MIT-67 Indoor Dataset. This dataset contains 67 different labels of indoor scenes.



The dataset used by us which contains only 10 classes with a total of 5059 images.



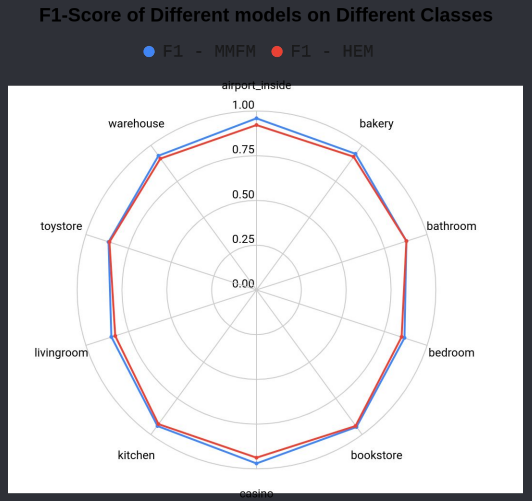
RESULTS

For MMFM,

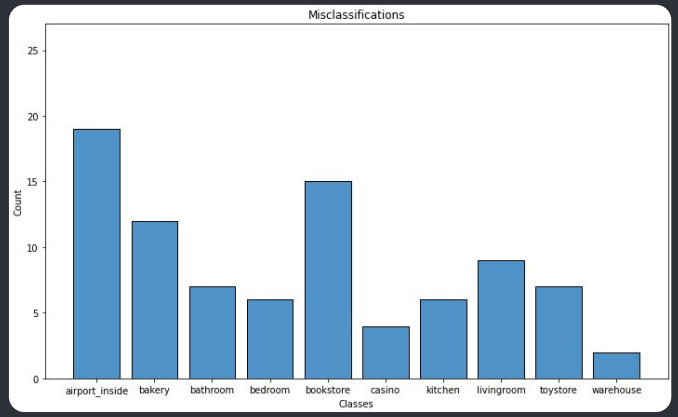
Model	Precision	Recall	f1-score
VGG-16	0.85	0.84	0.84
ResNet50	0.92	0.90	0.91
EfficentNet-B3	0.90	0.90	0.90
Fusion Model	0.92	0.91	0.91

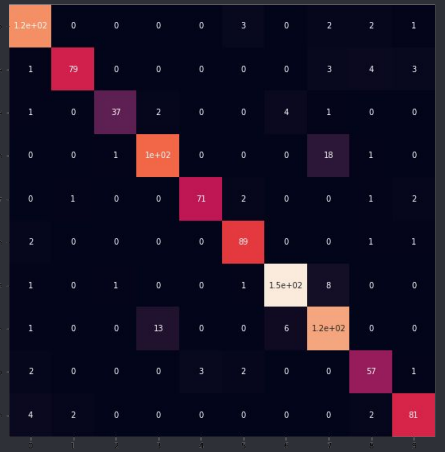
For HEM,

Model	Precision	Recall	f1-score
VGG-16	0.81	0.79	0.80
ResNet50	0.85	0.83	0.84
EfficentNet-B3	0.88	0.87	0.87
Fusion Model	0.90	0.89	0.90

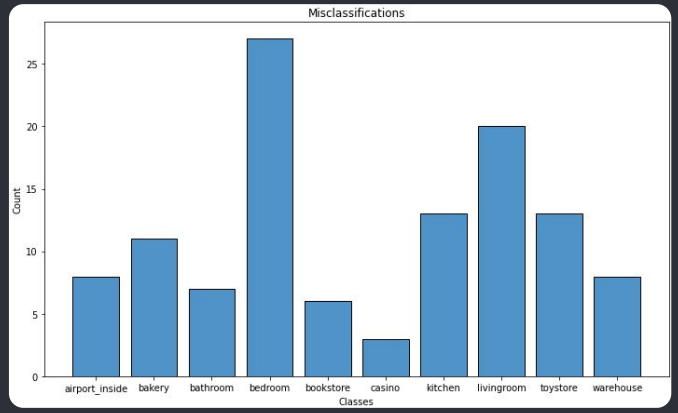


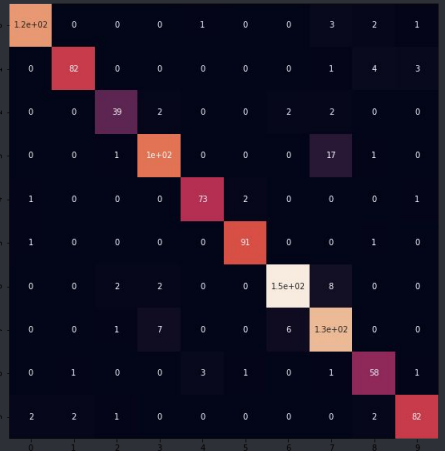
Multimodality Fusion Model





Hybrid Encoder Model





5 Software & Technology

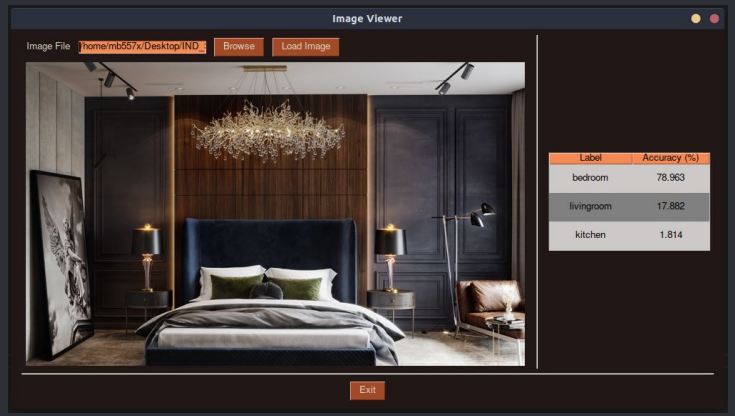
Software	Hardware
NumPy	Intel i7
Pandas	25 GB RAM
sklearn	NVIDIA GTX 1060
PySimpleGUI	
Keras	
Tensorflow	
Pickle	
Opencv	
Seaborn	

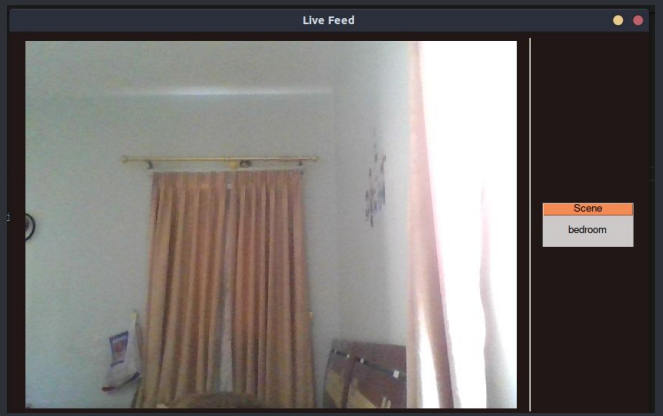
6 Graphical User Interface 

The ‘PySimpleGUI’ module in Python 3 was used for GUI creation.

Image Viewer

Live Feed





FUTURE SCOPE

- Indoor navigation for blind people.
- MRI based tumor detection.
- Activity recognition in Video sequences.
- This model is limited by the number of class labels, however work in the direction of increasing them to any possible scene is quite interesting and tangible.

Thank You.