AI HACKATHON 2019

Priority based audio alert system for visually impaired using pixel segmented data

Team Registration ID - AIH19T-0034

Participant Name	Organisation/ Institution	Email ID
Kunal Sinha	Pune Institute of Computer Technology	kunalsinha927@gmail.com
Bhushan Pagare	Pune Institute of Computer Technology	bpagare6@gmail.com
Kunal Chadha	Pune Institute of Computer Technology	kunalchadhaks2412@gmail.com
Aditya Chaloo	Pune Institute of Computer Technology	adichaloo@gmail.com

WHY THIS PROBLEM STATEMENT?

AI FOR HUMANITY

BUILD TO DEPLOY

PREVIOUS WORK

MY EYE 2.0

- OCR
- OBJECT DETECTION
- FACE DETECTION

MAPTIC

- WORN AS ACCESSORY AROUND NECK OR WRIST
- HAPTIC FEEDBACK

SEEING AI

- OCR
- OBJECT DETECTION
- FACE DETECTION
- SCENE DESCRIPTION
- FACE RECOGNITION

EYED

- OCR
- OBJECT DETECTION
- FACE DETECTION
- SCENE DESCRIPTION
- GEOLOCATION

NONE GIVE PRIORITY BASED ALERT

TO AID THE VISUALLY IMPAIRED

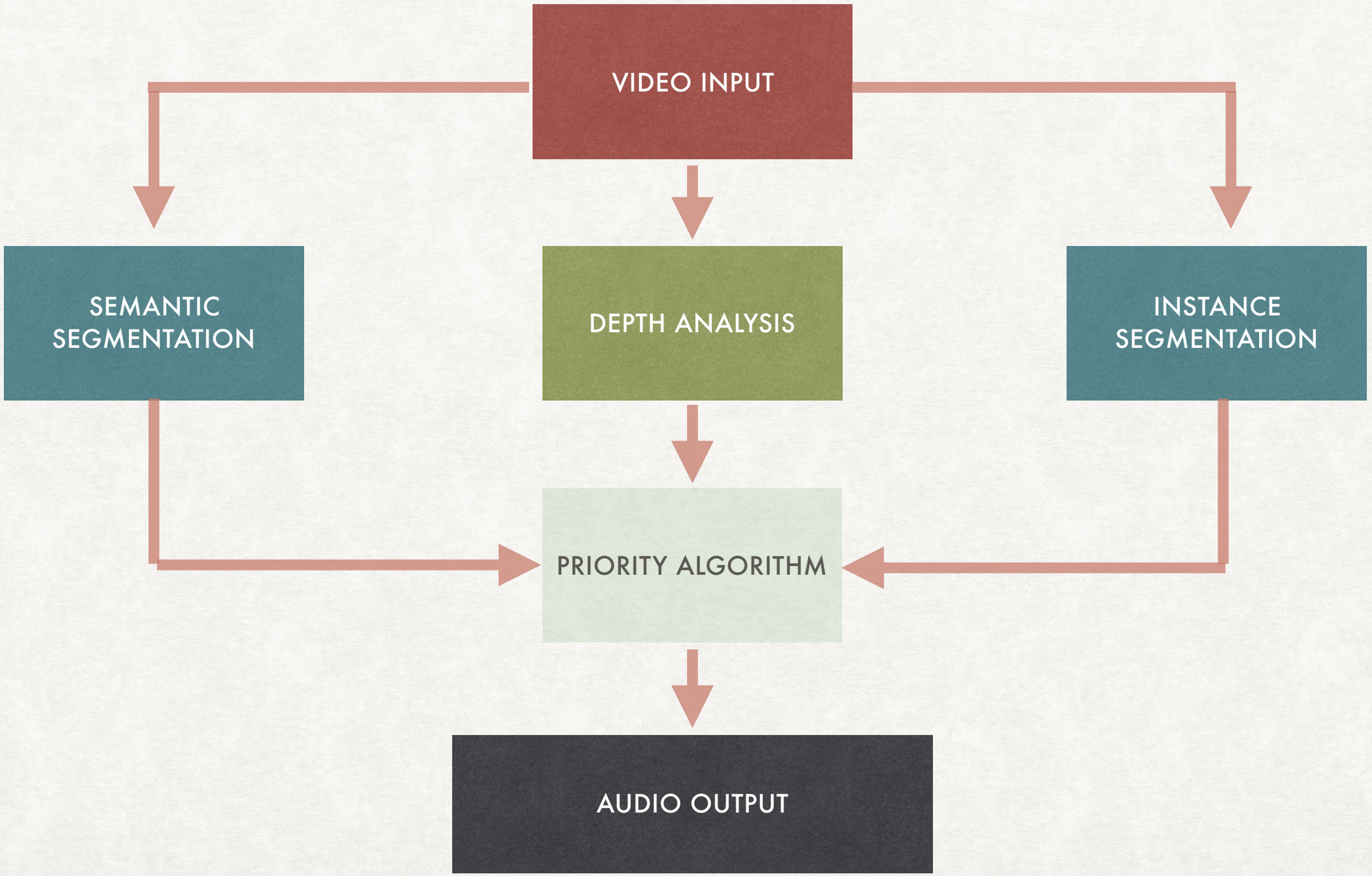
NAVIGATE IN AN UNKNOWN

ENVIRONMENT

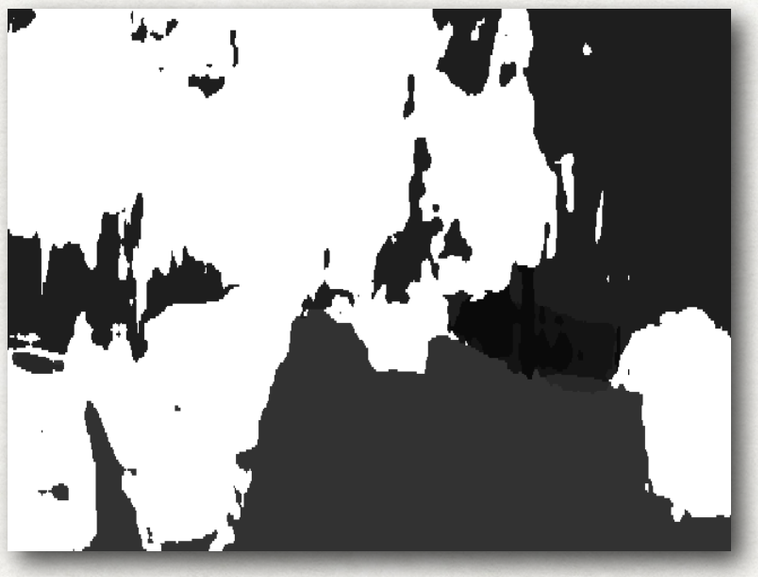
NETRA

State of the art application to aid visually impaired navigate in an unknown environment by giving priority based information, using deep learning and artificial intelligence.

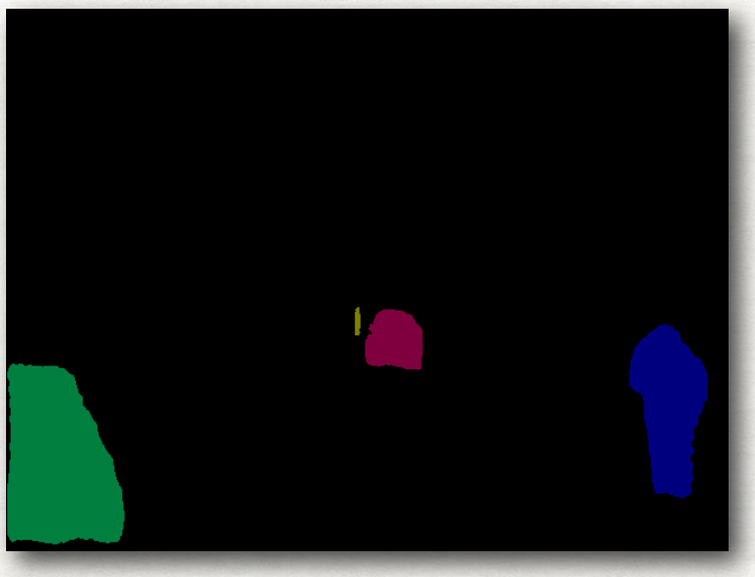
OUR SOLUTION



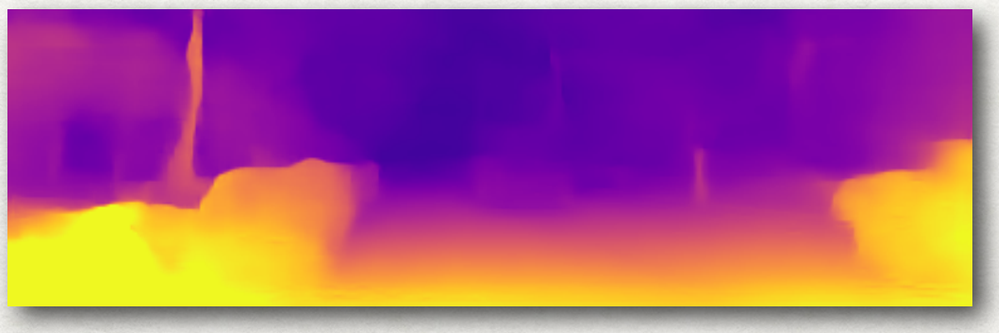
Semantic segmentation



Instance segmentation



Depth prediction



DETAILS OF OUR SYSTEM

SEMANTIC SEGMENTATION	DEPTH	INSTANCE SEGMENTATION
ESPNetv2	struct2depth	Mask RCNN (mobilenetv1 backbone)
2 classes: Road, Sidewalk	Monocular depth prediction	20 classes
A Light weight model to improve upon the inference time	State of the art algorithm for unsupervised monocular depth prediction for autonomous driving vehicles	State of the art algorithm for object detection
Future: Training on bigger dataset to improve IoU of road and side walk	Future: Model refinement on Indian street videos	Future: Transfer Learning on Indian Dataset

EXPERIMENTATION

DeeplabV3+

- Model **trained for 180 epochs** (took 9 hours)
- **Mean IoU: 67%**
- Average inference time: **0.6 secs**
- Total turnaround time: **1.3 secs**

*The time is calculated on Nvidia GTX 1070

Problems with it:

1. Model very heavy (1.06 GB); Cannot be deployed on any edge device.
2. Object motion cannot be incorporated (future scope).

Label_id	Label	IoU	Precision
0	Sky	91.21	0.94
1	Building	83.21	0.94
2	Column-pole	31.12	0.37
3	Road	94.90	0.98
4	Sidewalk	83.16	0.90
5	Tree	75.63	0.85
6	Sign-Symbol	47.63	0.53
7	Fence	26.34	0.37
8	Car	88.93	0.92
9	Pedestrian	56.32	0.68
10	Bicyclist	56.43	0.71

Class-wise IoU and Precision

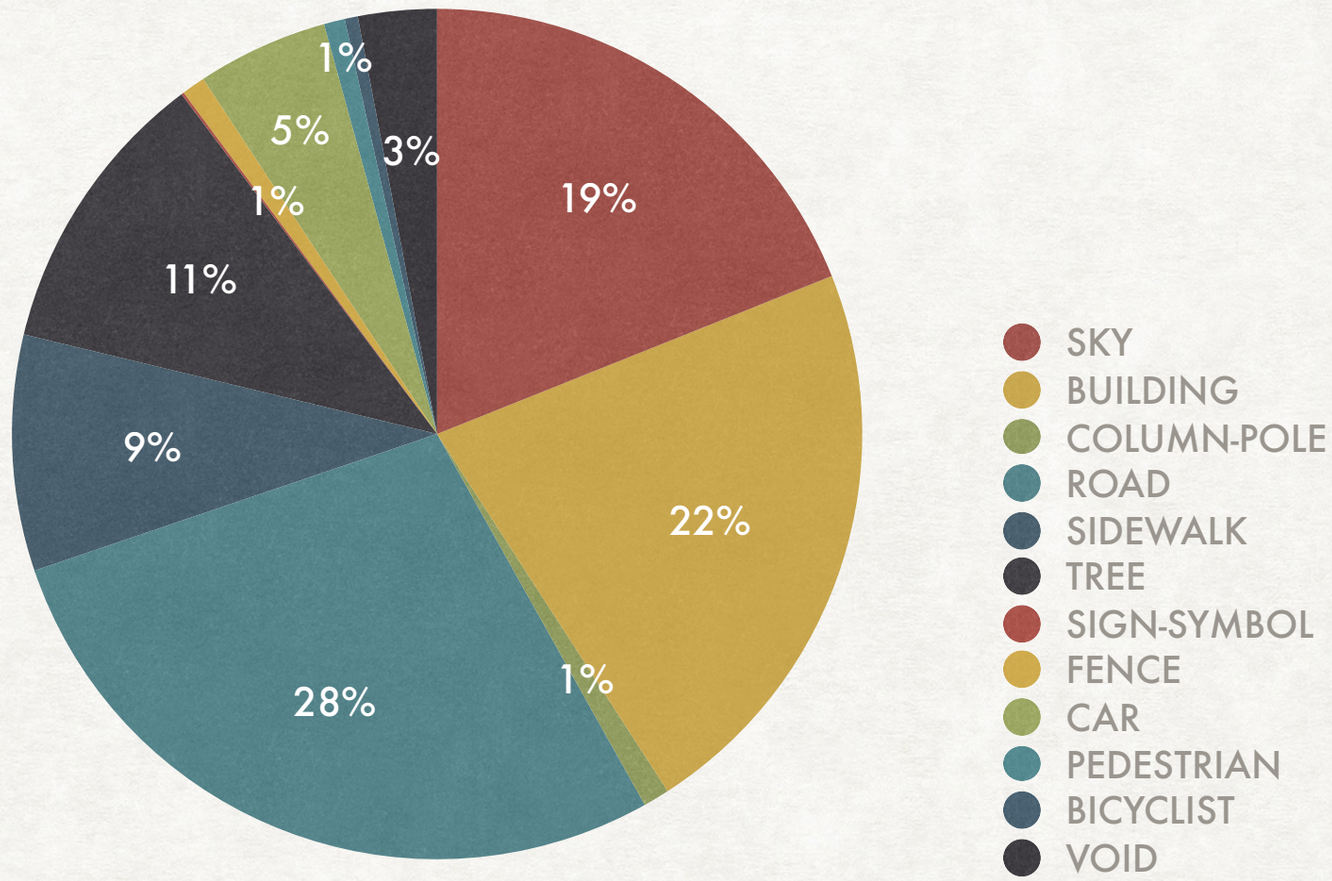
Jetson Nano

ESPNetV2	Struct2depth	Mask RCNN
0.20	0.23	1.7

Inference time on Jetson Nano
(sec per frame)

DATASET DETAILS

- Dataset for round 2: Cambridge-driving Labeled Video Database (CamVid).
- Consisted of 11 labeled classes.



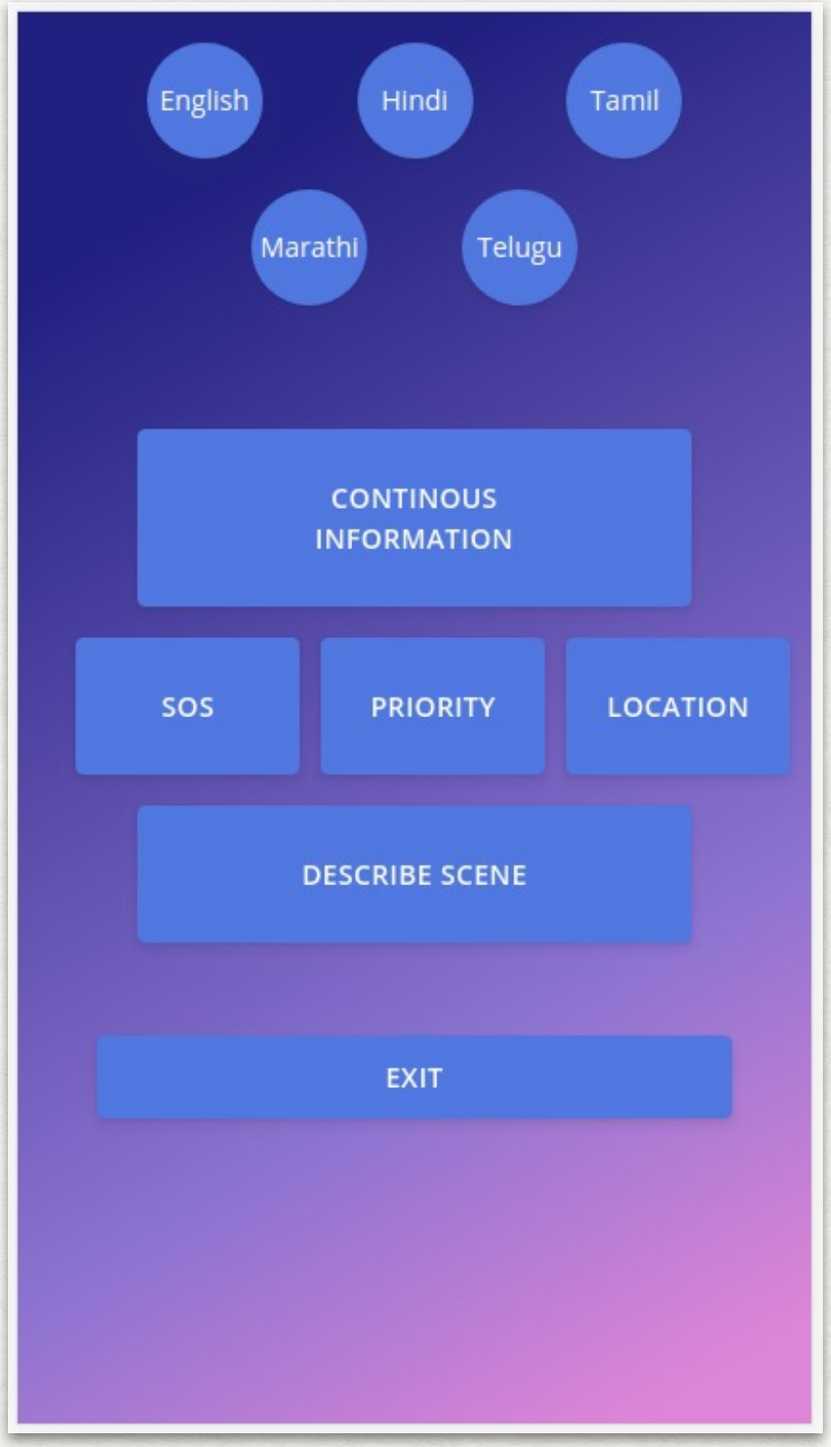
- Dataset for round 3: Images of Indian streets but only one labeled class - road.

RESULTS OF OUR SYSTEM

SEMANTIC SEGMENTATION (ESPNetv2)	DEPTH (struct2depth)	Mask RCNN (mobilenetv1 backbone)
mIoU on Camvid: 90.78%	Abs Rel: 0.1231 Sq Rel: 1.4367 RMSE: 5.3099 RMSE log: 0.2043	AP: 17.4
Inference time: 0.02 sec per frame	Inference time: 0.02 sec per frame	Inference time: 0.1 sec per frame

- Inference time of the integrated system is - **0.18 secs per frame (5 FPS)**
- Total turn around time - **0.6 secs per frame**

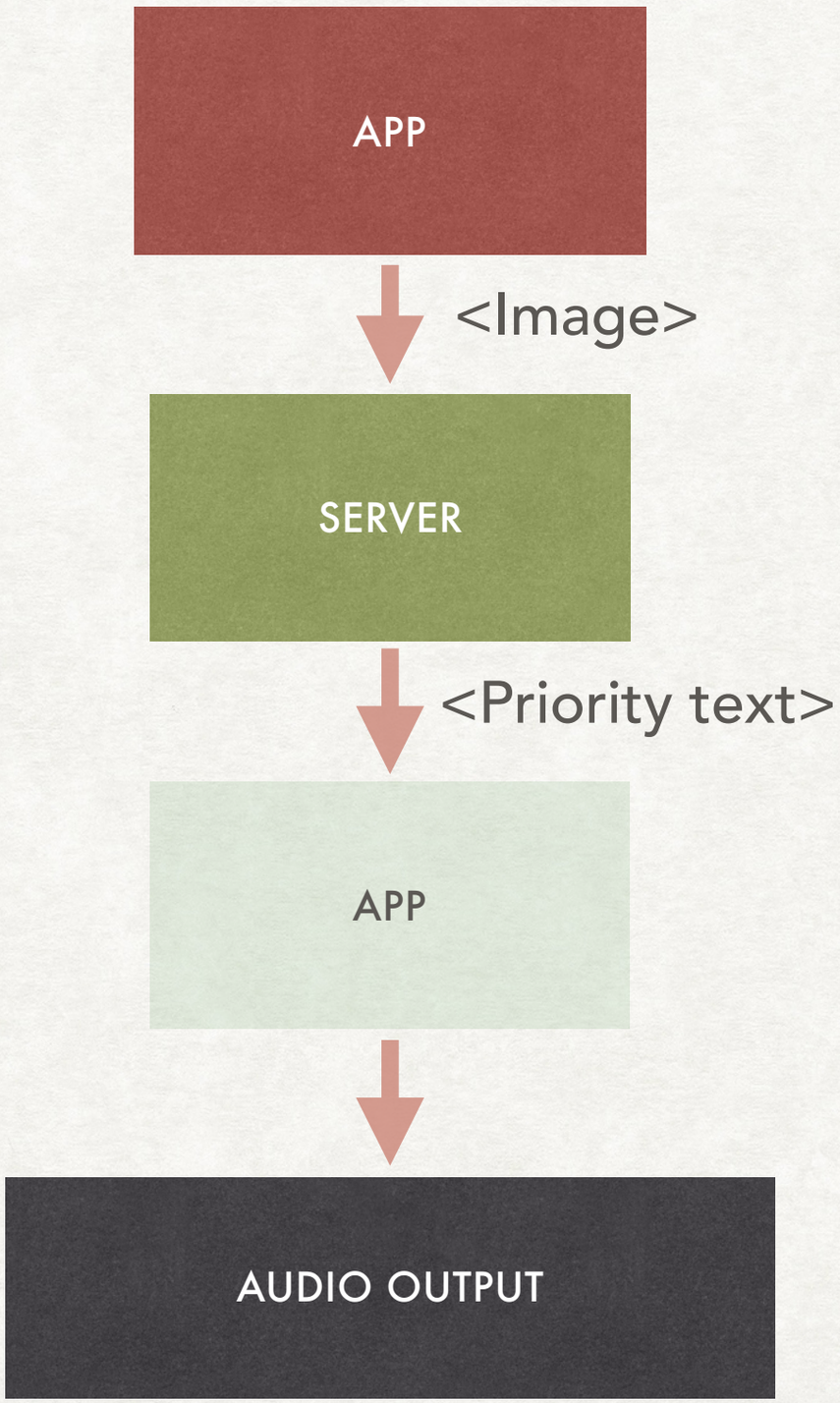
UI



Multi-lingual

Haptic feedback

DEPLOYMENT



FUTURE

- Training on Indian dataset with classes like rickshaw, bullock carts and regular Indian traffic.

- Deployment on an edge device.

THANK YOU

PATH TO FOLDER

- On Sreshtha: /home/aih15/AIH19T-0034