

CATALIST: CAmera TrAnsformations for multi- Lingual Scene Text recognition

<https://catalist-2021.github.io/>

Shivam Sood¹, Rohit Saluja², Ganesh Ramakrishnan¹, and Parag Chaudhuri¹

¹{ssood,ganesh,paragc}@cse.iitb.ac.in, ²rohit.saluja@research.iiit.ac.in

Contents

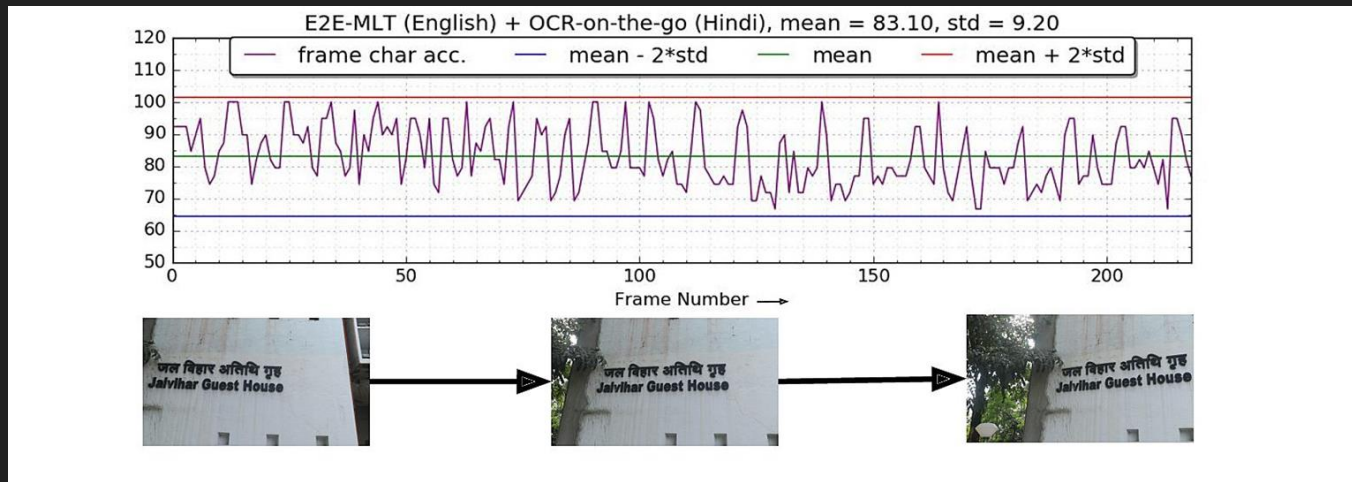
- Motivation
- CATALIST_d Dataset
- ALCHEMIST Synthesizer
- CATALIST Model
- Results

Motivation

Simple Bilingual Video Exhibiting Pan



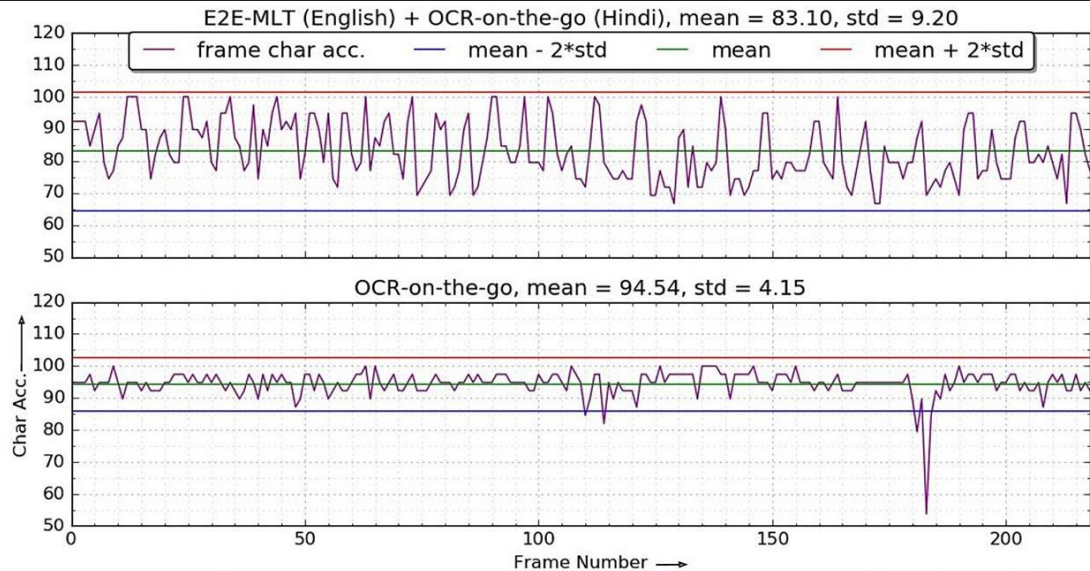
Results of Existing Models on Simple Video Exhibiting Pan



Buřta, Michal, Yash Patel, and Jiri Matas. "E2e-mlt-an unconstrained end-to-end method for multi-language scene text." In *Asian Conference on Computer Vision*, pp. 127-143. Springer, Cham, 2018.

Saluja, Rohit, Ayush Maheshwari, Ganesh Ramakrishnan, Parag Chaudhuri, and Mark Carman. "Ocr on-the-go: Robust end-to-end systems for reading license plates & street signs." In 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 154-159. IEEE, 2019.

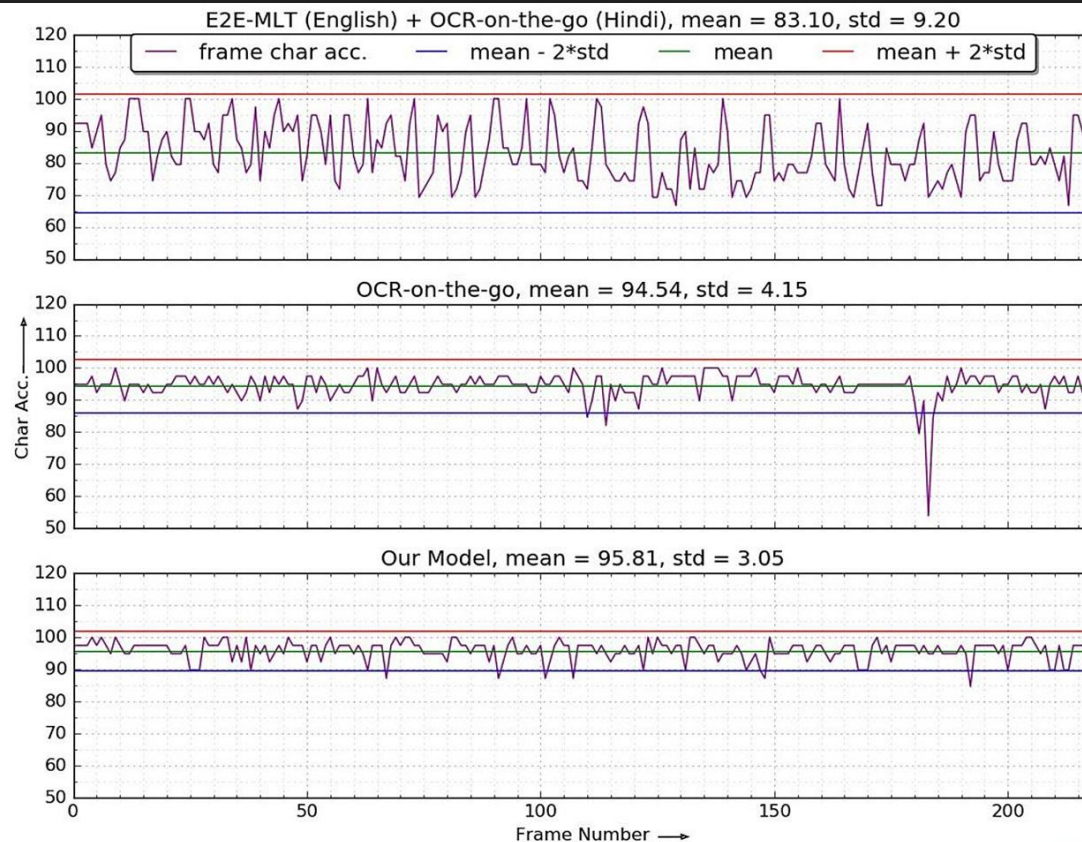
Results of Existing Models on Simple Video Exhibiting Pan



Bušta, Michal, Yash Patel, and Jiri Matas. "E2E-MLT - an Unconstrained End-to-End Method for Multi-Language Scene Text." In *Asian Conference on Computer Vision*, pp. 127-143. Springer, Cham, 2018.

Saluja, Rohit, Ayush Maheshwari, Ganesh Ramakrishnan, Parag Chaudhuri, and Mark Carman. "OCR On-the-Go: Robust End-to-end Systems for Reading License Plates & Street Signs" In 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 154-159. IEEE, 2019.

Our Model



CATALIST_d Dataset

CATALIST_d Setup



resolution: 1920x1080, fps: 25

CATALIST_d Samples



Zoom



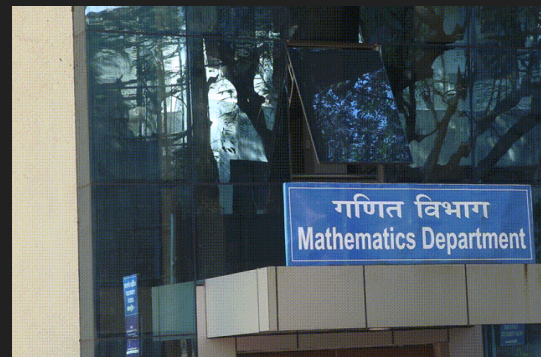
Roll



Tilt



Pan



Translate

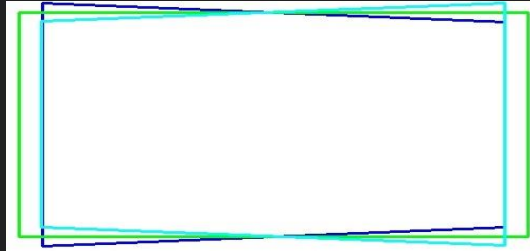
Distribution of Videos in CATALIST_d

| S.No. | Transformation Type | Number of Videos |
|--------------|----------------------------|-------------------------|
| 1 | Translation | 736 |
| 2 | Roll | 357 |
| 3 | Tilt | 387 |
| 4 | Pan | 427 |
| 5 | Zoom | 402 |
| Total | | 2309 |

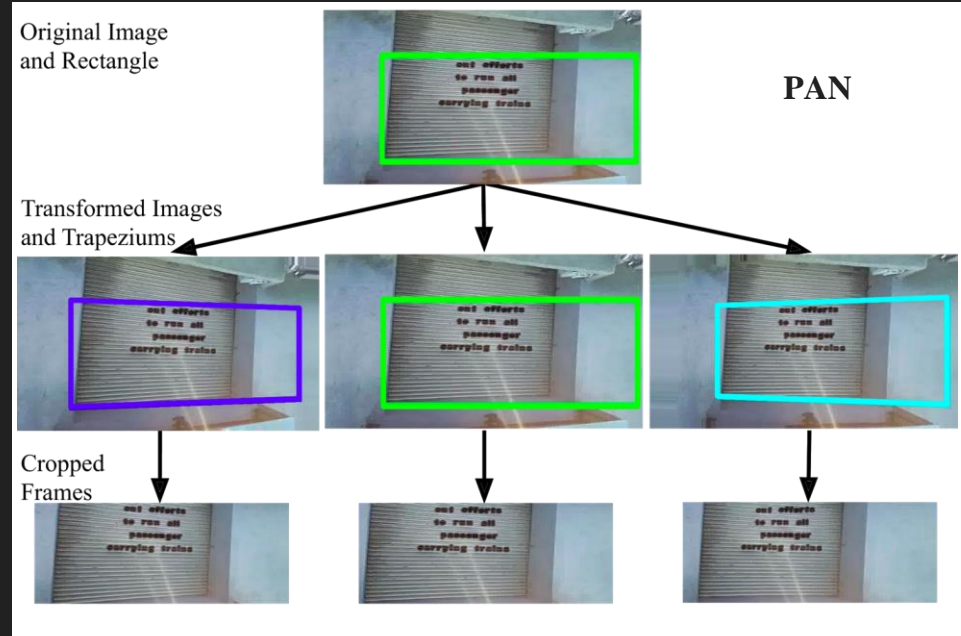
ALCHEMIST

Synthesizer

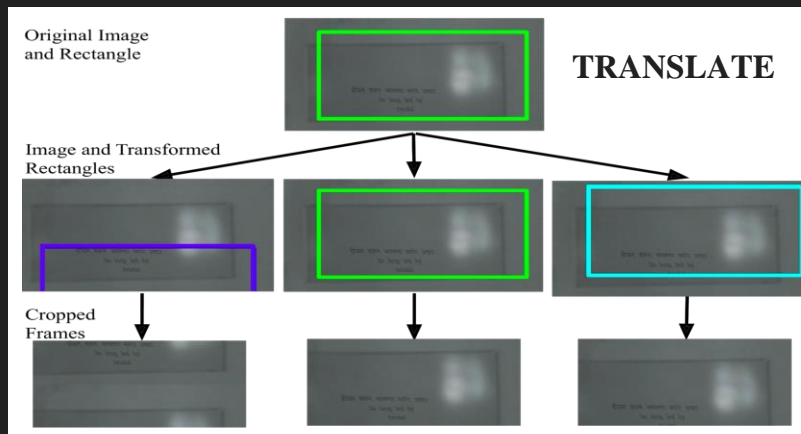
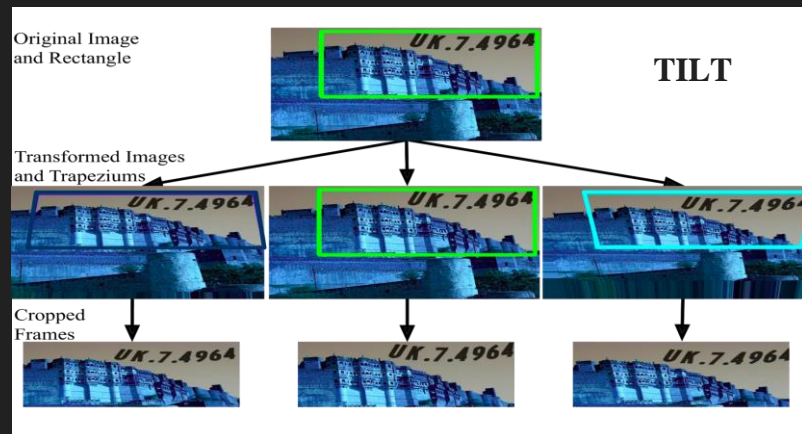
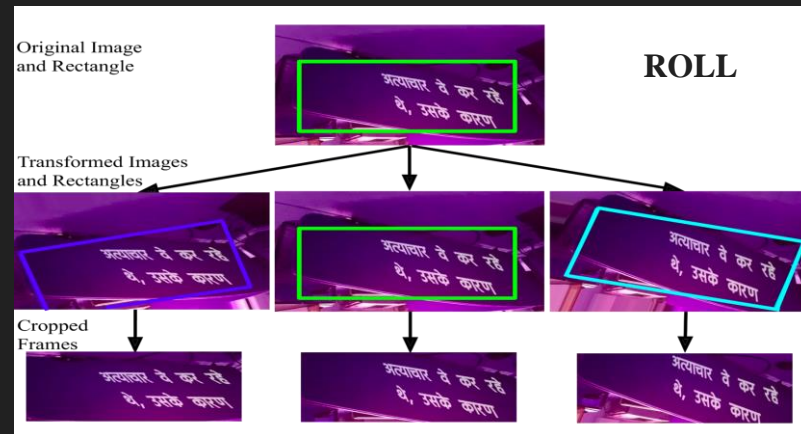
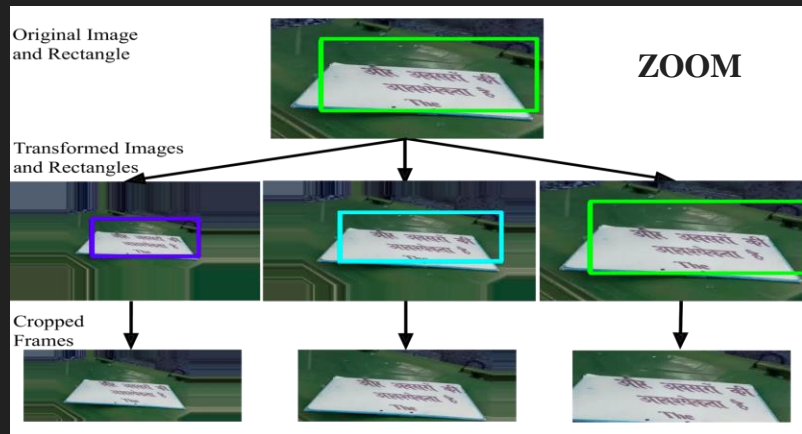
Homography Between Rectangle Corners & Equidistant Points



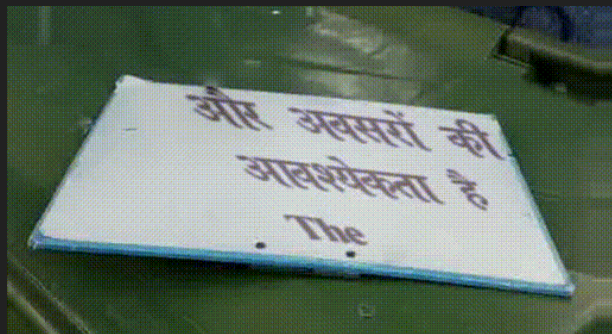
For PAN videos, we find Homography between rectangle corners (green) and 4 equidistant points (blue trapeziums)



Homography Between Rectangle Corners & Four Points



ALCHEMIST Samples (Total 12000 Videos: 5 sec. 25 fps)



Zoom



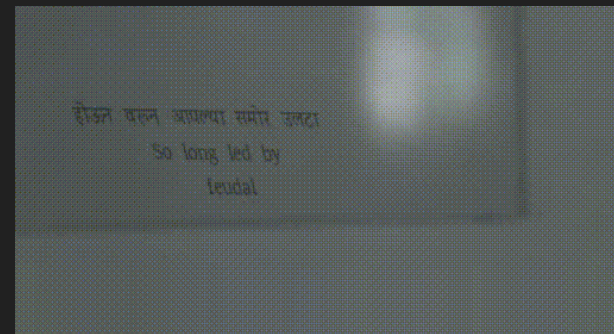
Roll



Tilt

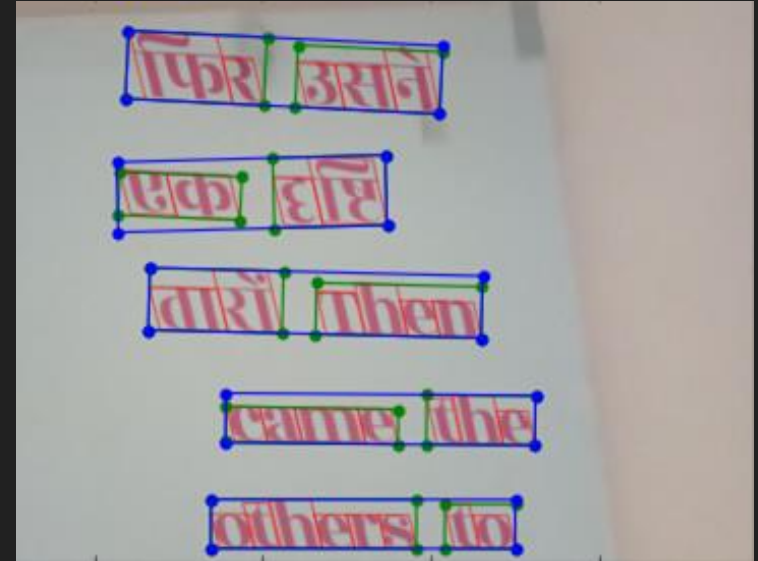
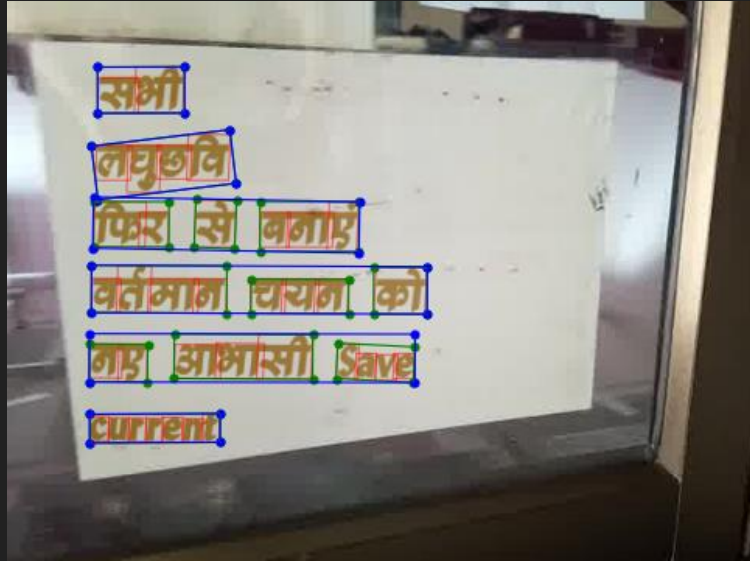


Pan



Translate

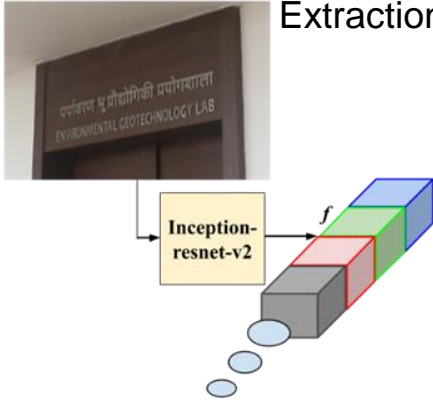
Bounding Boxes at Multiple levels for Supervision at each Frame



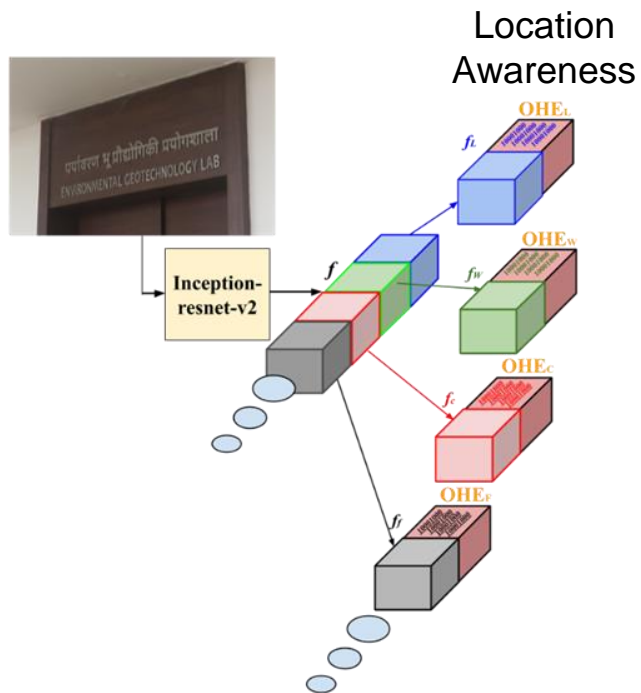
CATALIST Model

The CATALIST Model

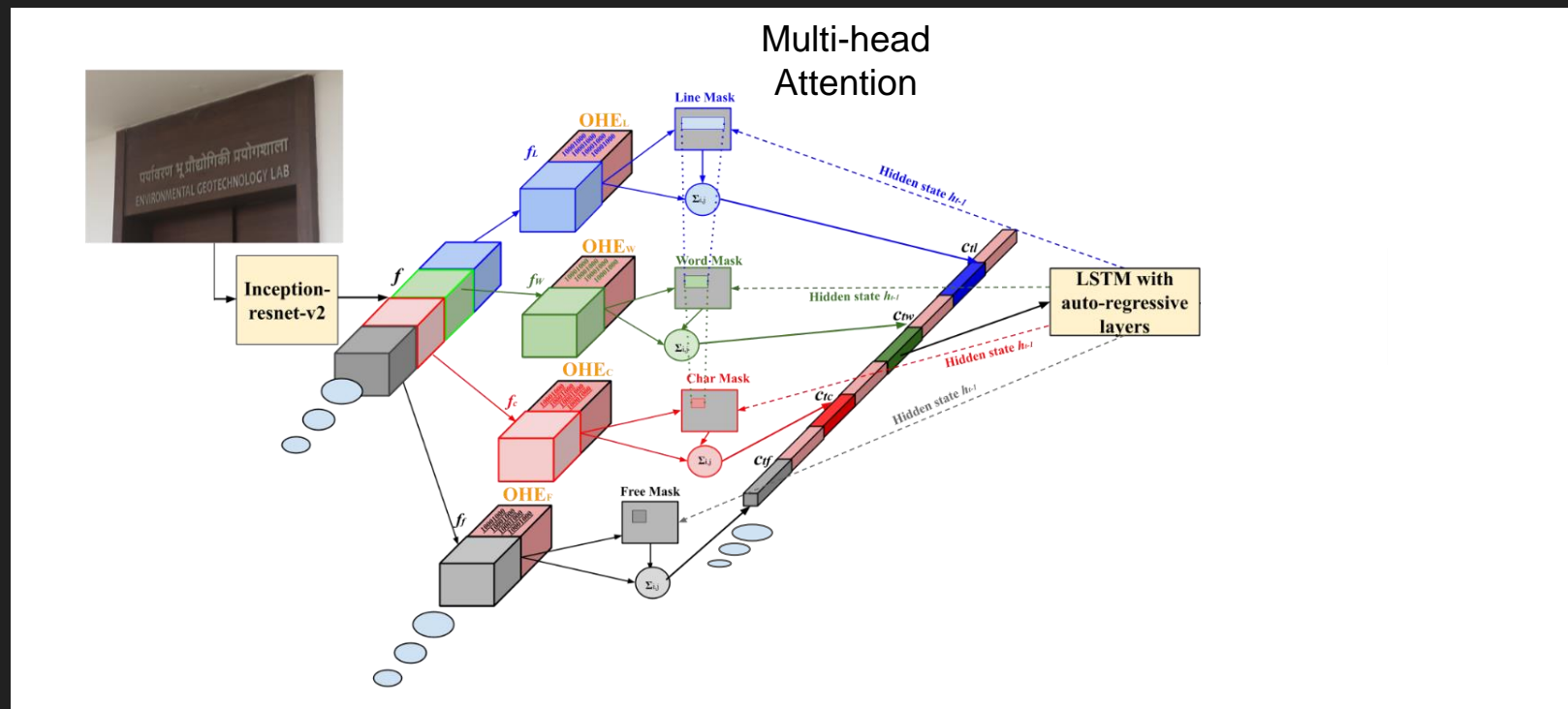
Feature
Extraction



The CATALIST Model

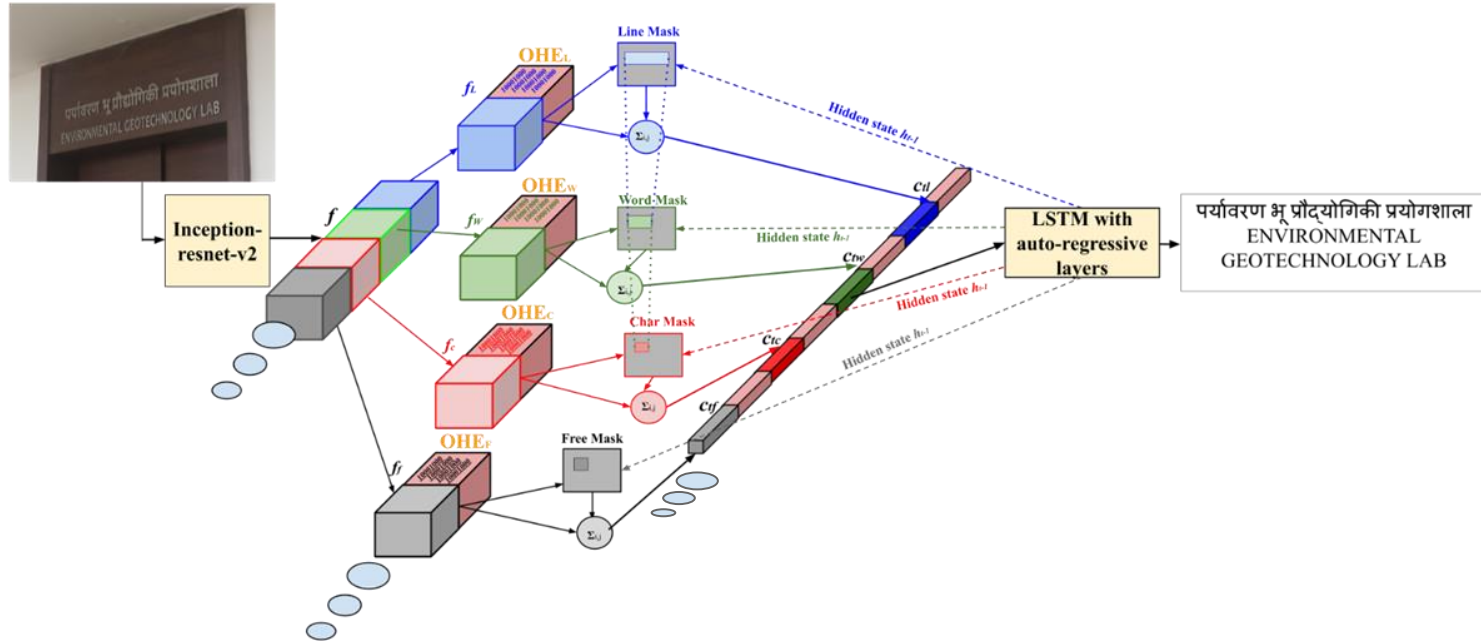


The CATALIST Model

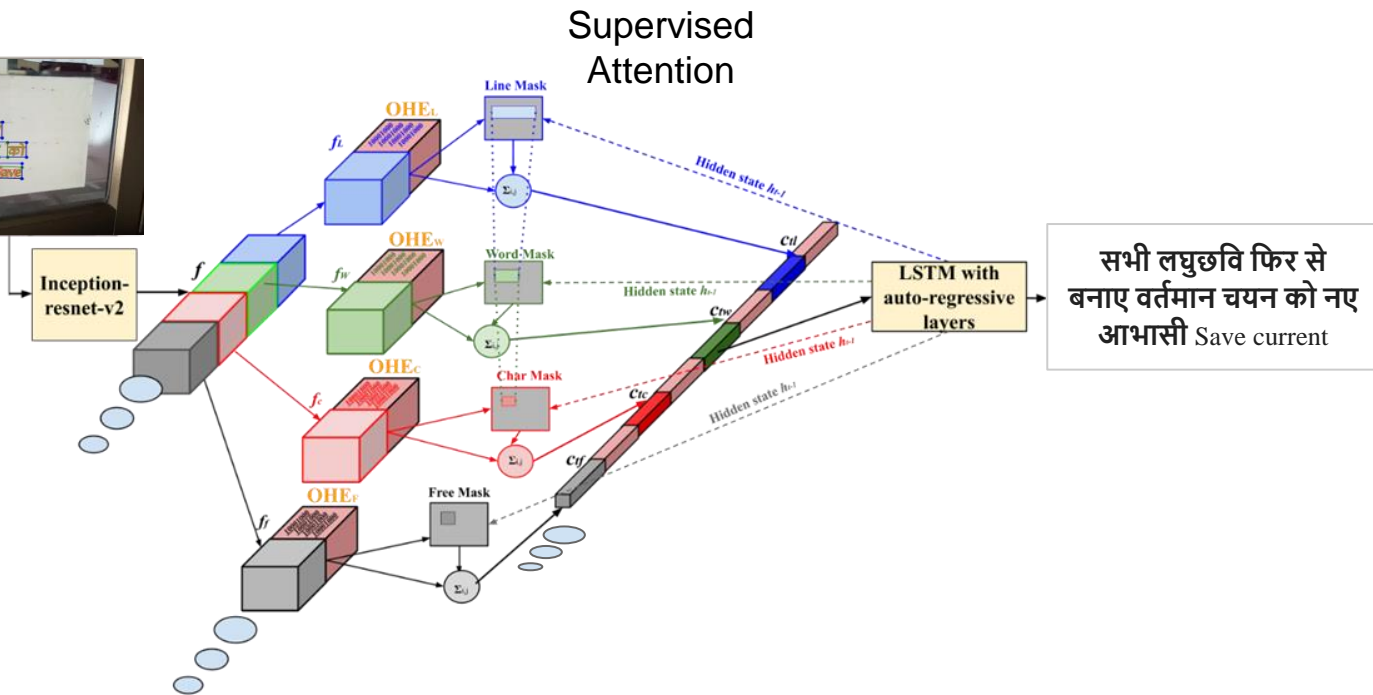


The CATALIST Model

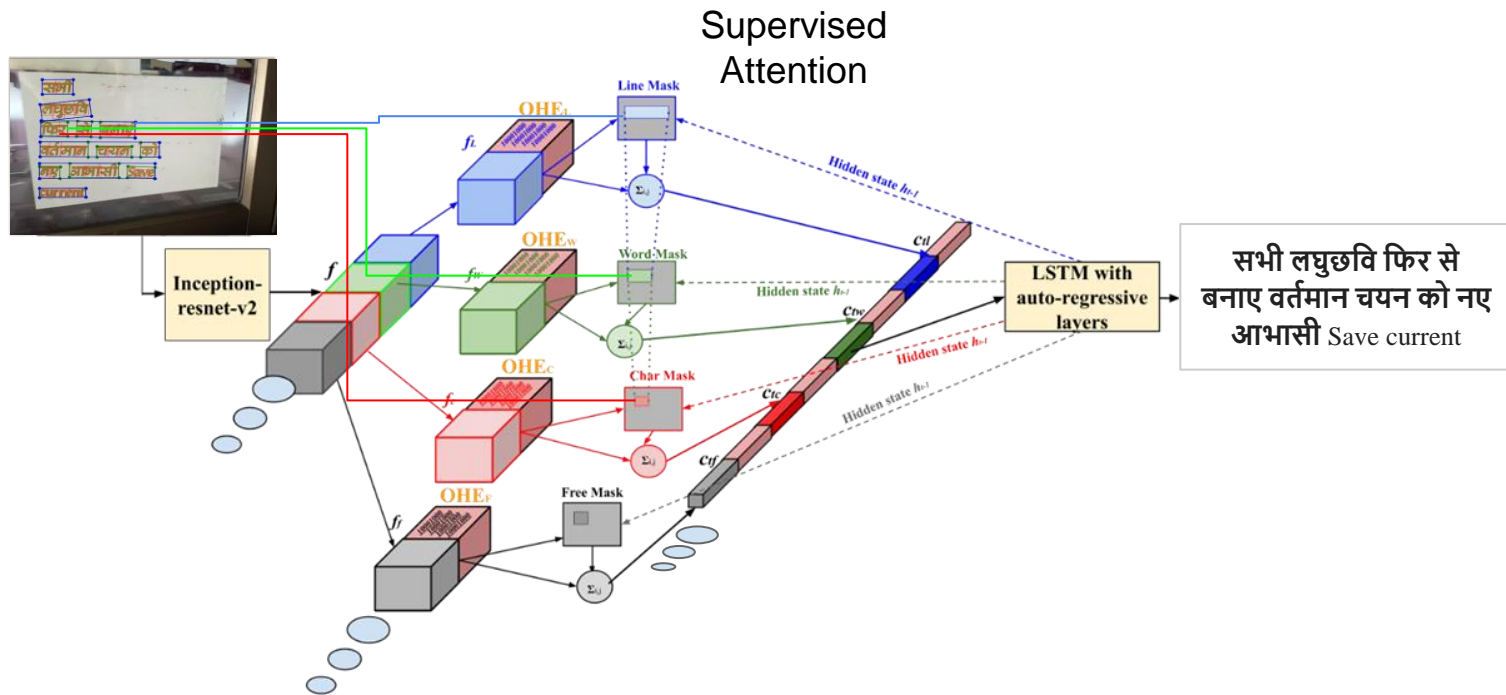
End-to-end Labels



The CATALIST Model



The CATALIST Model



Results

Results on Videos recorded with Cell Phone (Hand-held)

| S.No. | Training Model | Training Data | Test Data | Char. Acc. | Seq. Acc. |
|-------|----------------------------------|---------------|----------------------------------|------------|-----------|
| 1. | OCR-on-the-go (8 free masks) | OCR-on-the-go | OCR-on-the-go 200 test videos | 35.00 | 1.30 |
| 2. | CATALIST model (8 free masks) | CATALIST_ALL | | 65.50 | 7.76 |

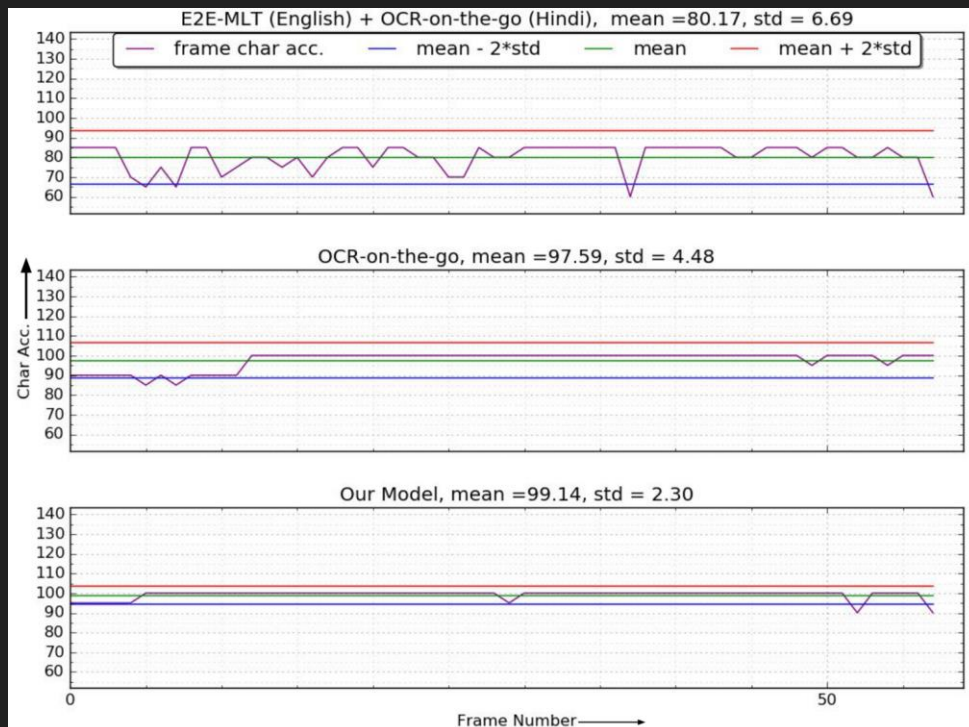
Results on Videos recorded with Cell Phone (Hand-held)

| S.No. | Training Model | Training Data | Test Data | Char. Acc. | Seq. Acc. |
|-------|---|---------------|----------------------------------|--------------|-------------|
| 1. | OCR-on-the-go (8 free masks) | OCR-on-the-go | OCR-on-the-go 200 test videos | 35.00 | 1.30 |
| 2. | CATALIST model (8 free masks) | CATALIST_ALL | | 65.50 | 7.76 |
| 3. | CATALIST model (3 superv., 5 free masks) | CATALIST_ALL | | 68.67 | 7.91 |

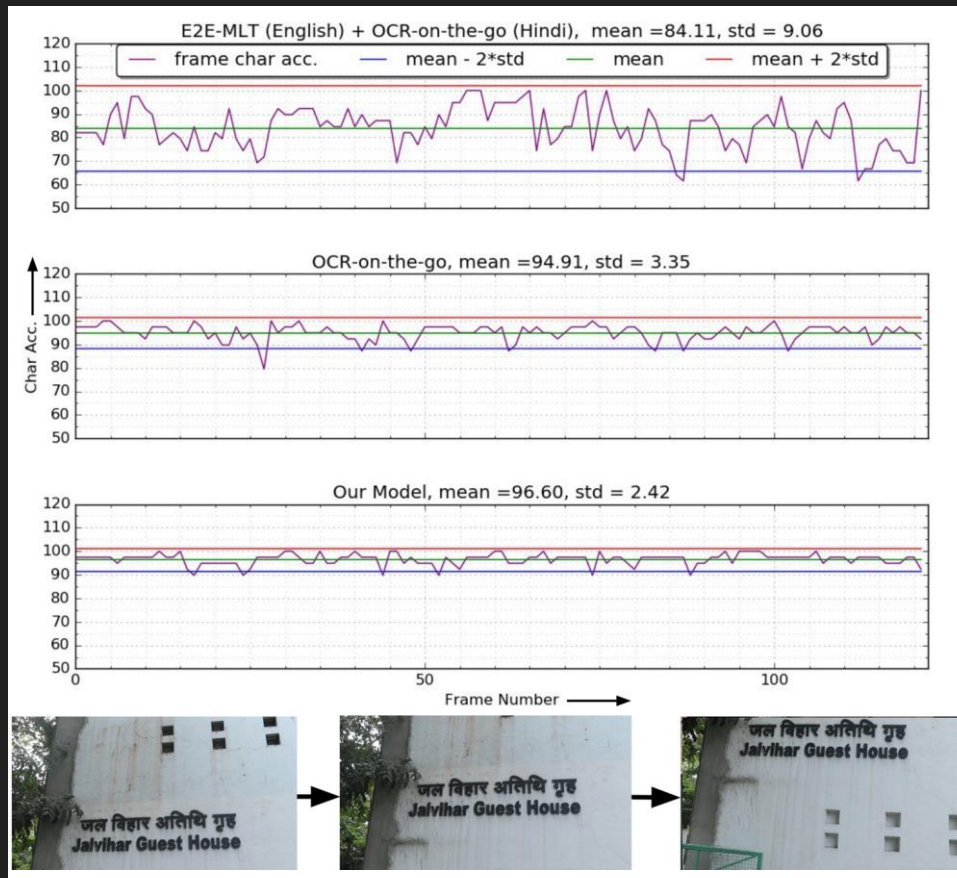
Results on CATALIST_d (recorded with Camera on Tripod)

| S.No. | Training Model | Training Data | Test Data | Char. Acc. | Seq. Acc. |
|-------|---|---------------|-------------------------------------|--------------|-------------|
| 1. | OCR-on-the-go (8 free masks) | OCR-on-the-go | OCR-on-the-go 200 test videos | 35.00 | 1.30 |
| 2. | CATALIST model (8 free masks) | CATALIST_ALL | | 65.50 | 7.76 |
| 3. | CATALIST model (3 superv., 5 free masks) | CATALIST_ALL | | 68.67 | 7.91 |
| 4. | CATALIST model (8 free masks) | CATALIST_ALL | 491 CATALIST _d videos | 73.97 | 6.50 |
| 5. | CATALIST model (3 superv., 5 free masks) | | | 73.60 | 7.96 |

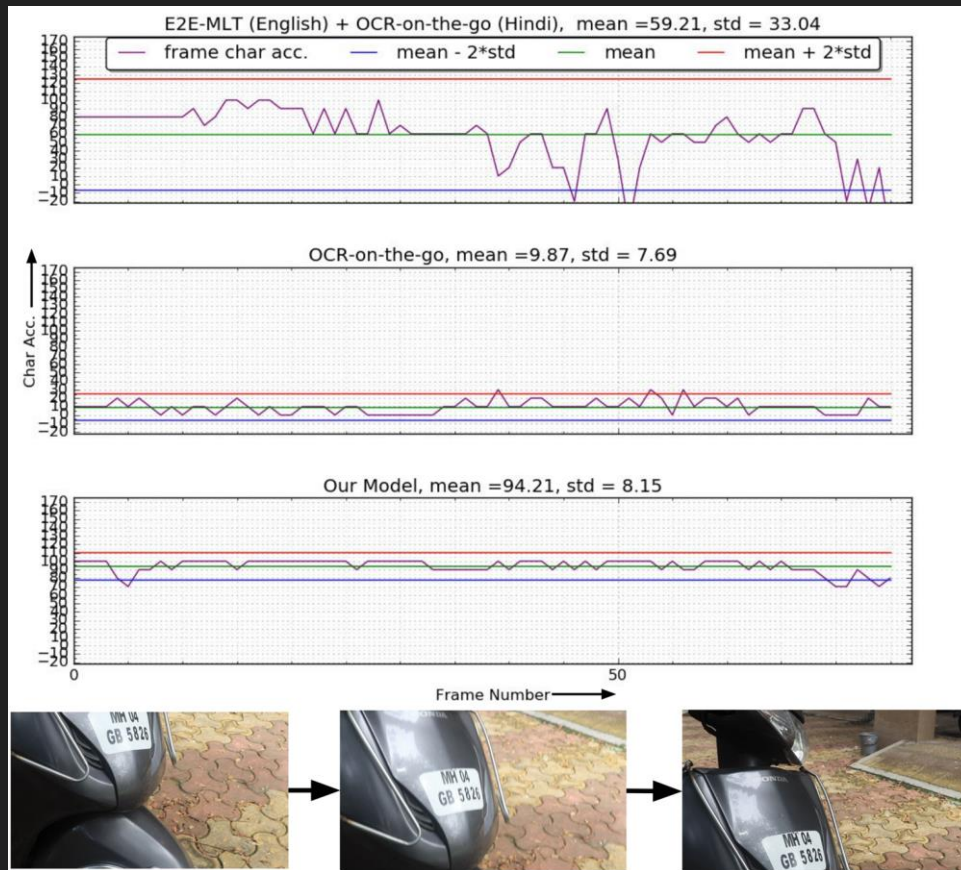
Qualitative Results



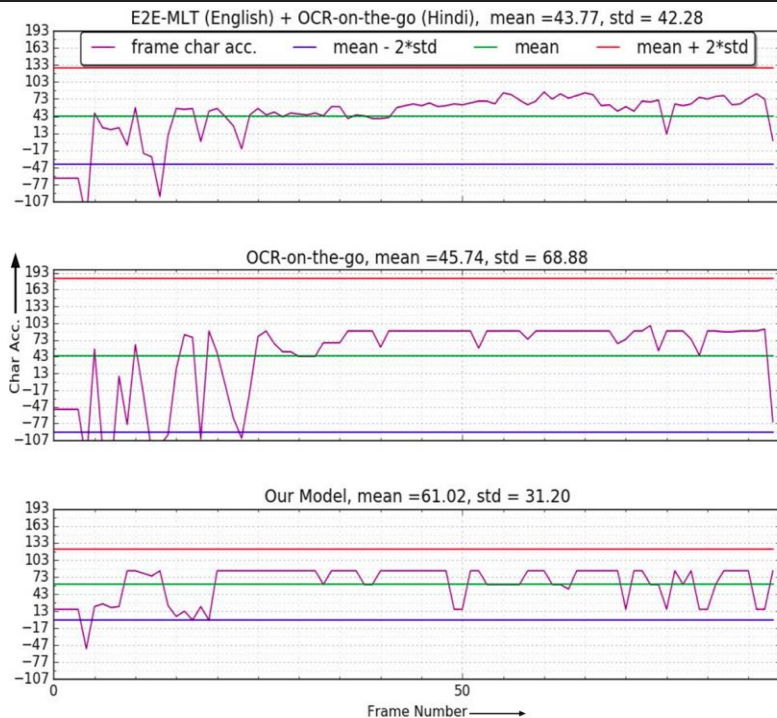
Qualitative Results



Qualitative Results



Qualitative Results



Thank You

<https://catalist-2021.github.io/>

Shivam Sood¹, Rohit Saluja², Ganesh Ramakrishnan¹, and Parag Chaudhuri¹

¹{ssood,ganesh,parag}@cse.iitb.ac.in, ²rohit.saluja@research.iiit.ac.in