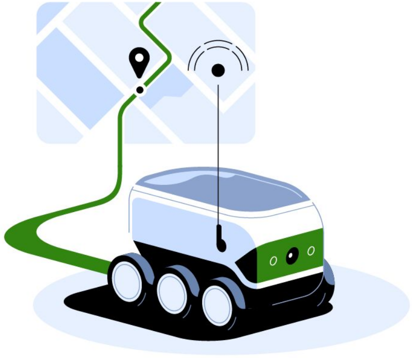
CMP 630 - Project

Submitted to: Dr. Abdelmoniem bayoumi
Submitted by: Eng. Mahmoud Saad

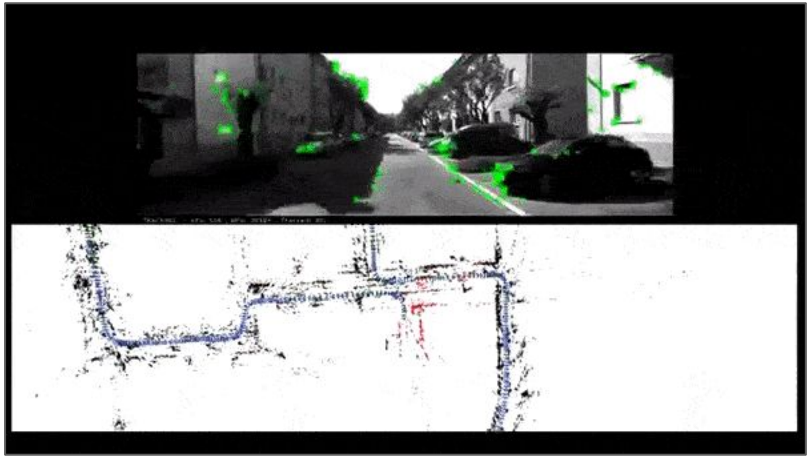
Agenda

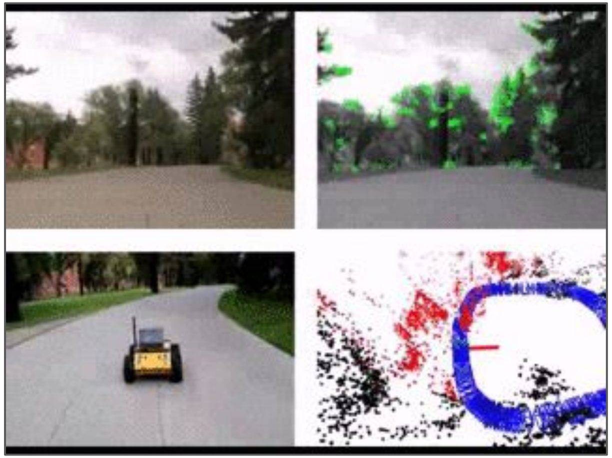
1. SLAM & Autonomous Navigation robots
2. Loop closure detection problem
    a. Explanation of the problem
    b. Benefits of the problem
3. Existed solutions
4. Proposed approach
    a. Large models
    b. SAM: Segment Any thing model
5. Dataset and Experiments
6. Results and Conclusion
7. Future work



SLAM & Autonomous Navigation robots

is **a method to create a map of the environment while tracking the position of the map creator**. This mapping and positioning method is the key piece in enabling robots to autonomously know their current location in space and navigate to a new location





SLAM & Loop closure

For self-driving cars to localize themselves in a global environment, accurate geometric maps are needed. Self-driving cars need to know in which lane they are driving but also how far away from the lane boundary they are. The accuracy needs to be on the order of 5 cm.

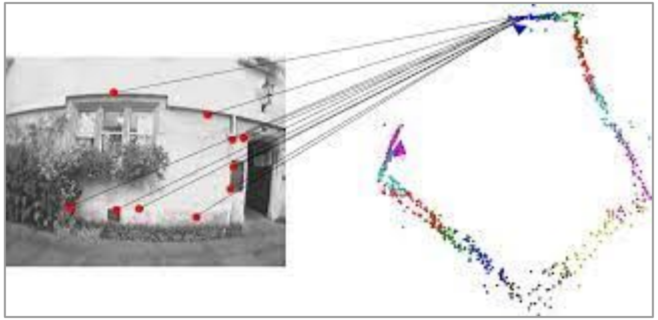
Structure from Motion (SfM) and Simultaneous Localization and Mapping (SLAM)-based methods have been known for years for extracting a 3D environment from a series of images. However, these algorithms perform well on local but not on global scales. The odometry is extracted from the difference of the current image to the previous image(s). This leads to an incremental change in the camera pose and evidently to a drift over the measurement time. Oftentimes slow motion or just rotations of the camera can worsen the problem.

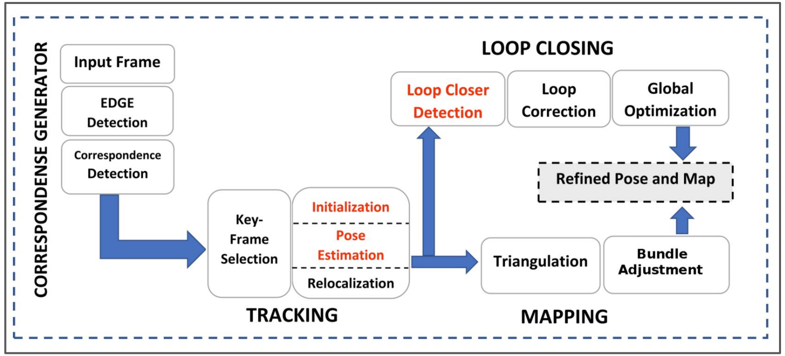
It is troublesome if a section 100 m down the road is mapped with a drift of several meters. Yes, localisation in a local environment still works well but this local section cannot be connected to other mapping campaigns. It can also lead to bad first-time localisation if I choose my starting point by an absolute position measurement such as the less accurate known GPS position.

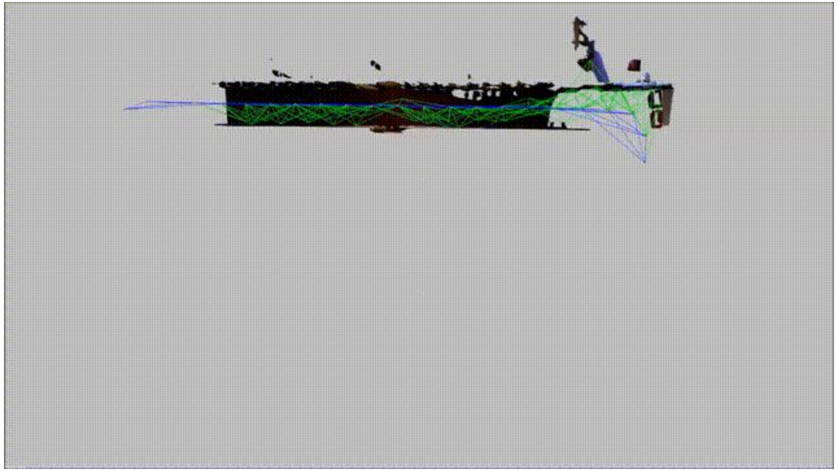
Loop closure detection

-Minimizing the localization error.

-Fix drifting issue







Existed Approaches

1. Traditional Techniques
2. Graph based approaches
3. CNN approaches
4. Vision Transformer approaches

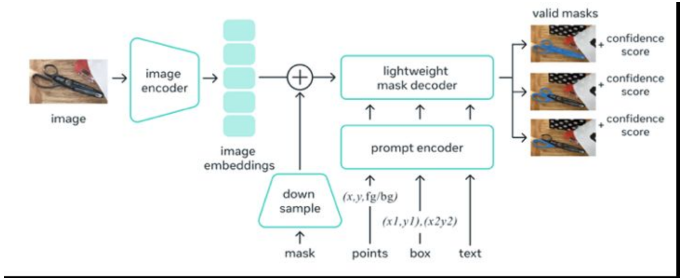
Q: are those methods are generalized?

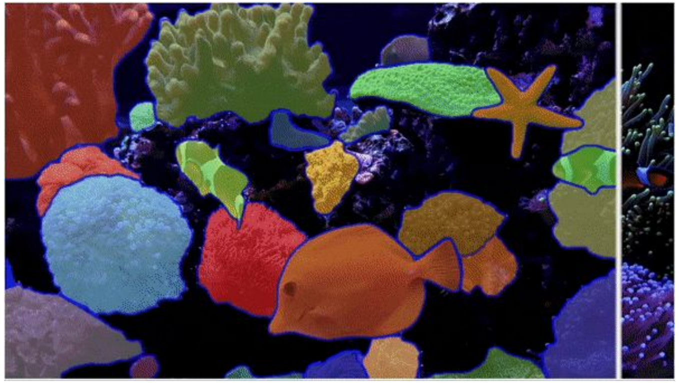
Proposed Approach

Large Language model

SAM Segment Anything model

Developed by FAIR

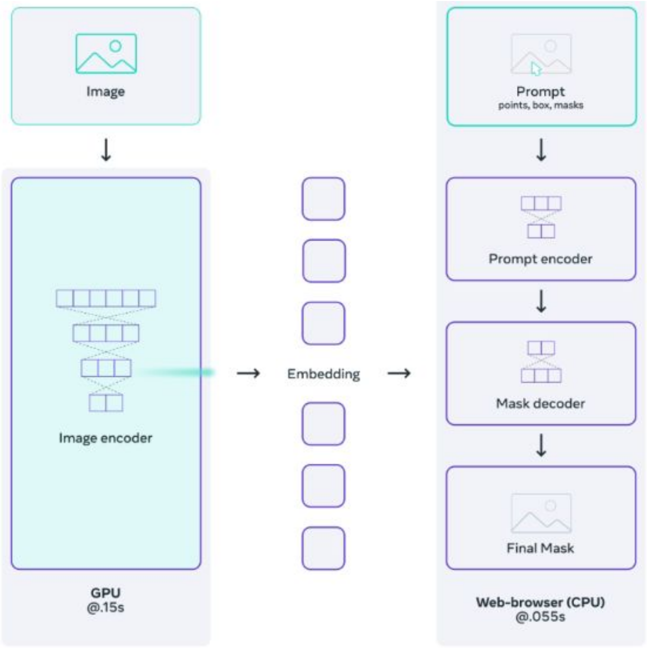




SAM

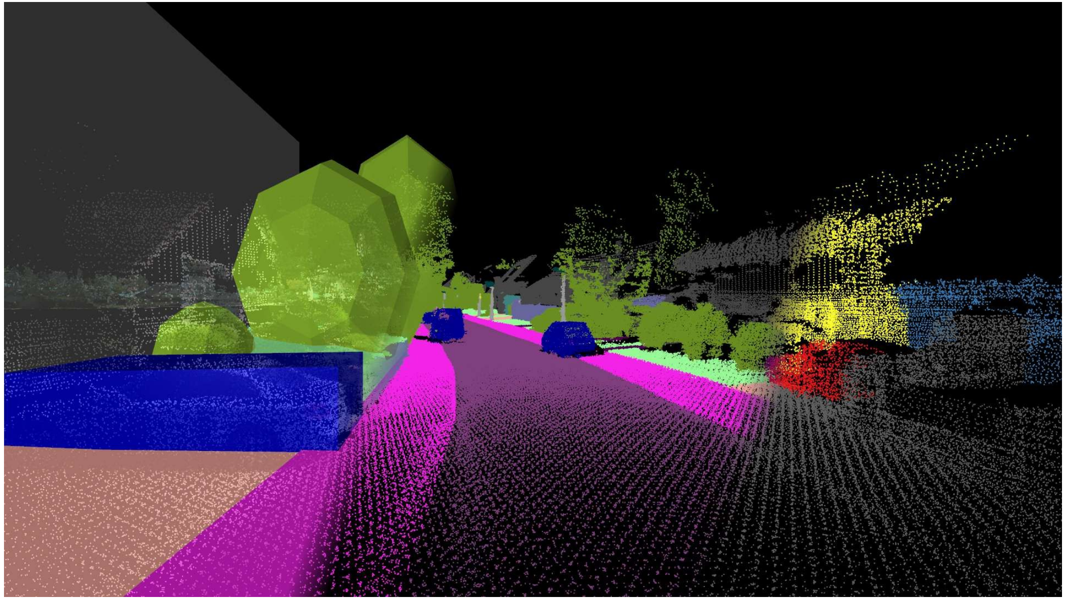
Encoder + Decoder (ViT)

final dataset includes more than 1.1 billion segmentation masks collected on ~11 million image



Dataset & Experiments

KITTI dataset Sequence 00 + 02 + 05



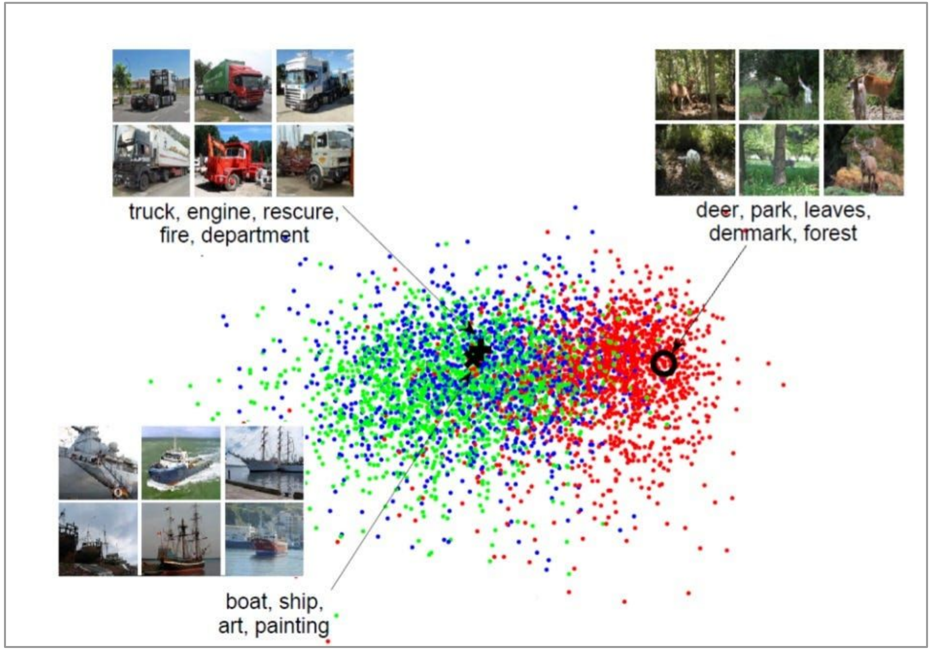
Results

Dataset	Approaches	Precision (%)	Recall Rate (%)
sequence 00	Kim [21]	100	87
	GLAROT3D [17]	86	40
	Cieslewski [24]	92	80
	Gálvez-López [10]	100	92
	Proposed	**100**	**90.2**
sequence 02	Kim [21]	90	73
	Gálvez-López [10]	100	80.6
	Proposed	**98**	**91**
sequence 05	Kim [21]	100	90
	Cieslewski [24]	93	60
	GLAROT3D [17]	80	80
	Gálvez-López [10]	100	87.6
	Proposed	**100**	**91.2**

Proposed SAM

Seq 05	Thres = 0.85	75	**99.2**
	Thres = 0.9	**100**	40

Results analysis



Future Work

Image description	Similarity
Encoder only	Cosine similarity
Encoder + decoder	Cosine similarity + Segmentation masks
Encoder + prompting Encoder + decoder	Cosine similarity + Segmentation masks + key matching similarity

Conclusion

- SAM is powerful technique for image understanding and can play role in generalization for any scene while Loop closure
- More work is need to explore all capabilities of SAM to serve LCD problem