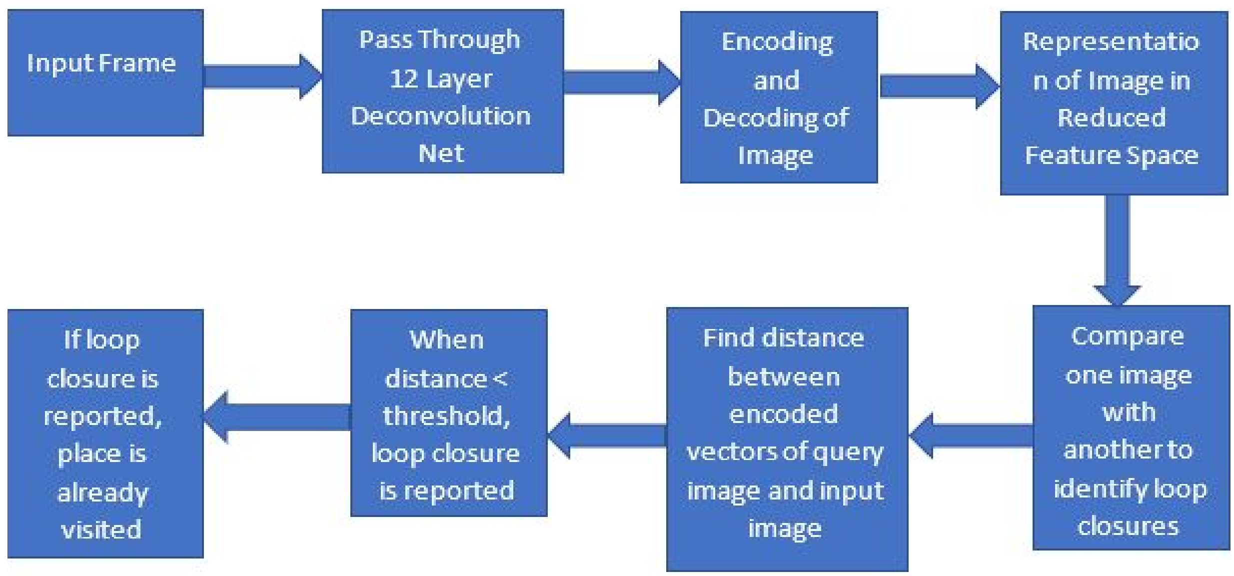
Learning Deep Representation for Place Recognition in SLAM

Understanding of Research Paper / Project

Place recognition is one of the most fundamental topics in computer vision and robotics communities, where the task is to accurately and efficiently recognize the location of a given query image.
It is to be noted that in the context of this work, place recognition refers to recognising whether a place has been visited previously or not.
In navigation, robotic mapping and odometry for virtual reality or augmented reality, **simultaneous localization and mapping (SLAM)** is the computational problem of constructing or updating a map of an unknown environment while simultaneously keeping track of an agent's location within it.
In the context of robot navigation with vision, the task of Simultaneous Localization and Mapping (SLAM) is an important task.
The entire SLAM process relies on recognizing the places the robot has already visited to achieve visual loop closure detection.
Closing loops for pose graph optimization, by recognizing previously mapped places is an essential step for performing SLAM.
The major tasks are like representing the frames with the help of visual descriptors and subsequently judging the similarity between the frames based on the descriptors.
An LCA is a fully connected 2-layer feedforward neural network where the number of input neurons is equal to the number of output neurons and the number of hidden neurons is equal to the dimension of the autoencoder, which in this case is 40.

Scope of Project

The methodology that we will be following for the project is as shown in the flowchart below:



Methodology

The architecture of the 12-layer deconvolution net will be replicated according to the one proposed in the research paper, which is as below:

Layer	Kernel size	Stride	Pad	Output dim.
Input	–	–	–	$1 \times 96 \times 336$
Conv-1	3×3	1	1	$2 \times 96 \times 336$
Conv-2	3×3	1	1	$3 \times 96 \times 336$
Pool-1	2×2	2	0	$3 \times 48 \times 168$
Conv-3	3×3	1	1	$5 \times 48 \times 168$
Conv-4	3×3	1	1	$8 \times 48 \times 168$
Pool-2	2×2	2	0	$8 \times 24 \times 84$
Conv-5	3×3	1	1	$5 \times 24 \times 84$
Pool-3	2×2	2	0	$5 \times 12 \times 42$
LCA-enc	–	–	–	5×40
LCA-dec	–	–	–	$5 \times 12 \times 42$
Unpool-1	2×2	2	0	$5 \times 24 \times 84$
Deconv-1	3×3	1	1	$8 \times 24 \times 84$
Unpool-2	2×2	2	0	$8 \times 48 \times 168$
Deconv-2	3×3	1	1	$5 \times 48 \times 168$
Deconv-3	3×3	1	1	$3 \times 48 \times 168$
Unpool-3	2×2	2	0	$3 \times 96 \times 336$
Deconv-4	3×3	1	1	$2 \times 96 \times 336$
Deconv-5	3×3	1	1	$1 \times 96 \times 336$

At the heart of the proposed architecture lies a deconvolution net. It is further modified by adding a layer of locally connected autoencoders to map an image frame into a representation vector of n dimensions. This will be developed using the functions available in the TensorFlow library.

Individual Responsibilities

- Ajinkya Bedekar: Develop the deconvolution net according to the proposed architecture
- Devansh Anhal: Training the model using unsupervised learning approach
- Dhruva Agarwal: Testing the model to check how effective the model is
- Harsha Deuri: Algorithm for Loop Closure Detection