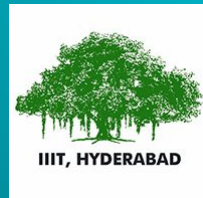


Improved Visual Relocalization by Discovering Anchor Points

Soham Saha

Girish Varma

C.V. Jawahar



The Visual Relocalization Problem

- Given an image, predict 6-DOF of camera. That is
 - location [3 spatial coordinates]
 - pose [3 angles].
- Needed in Autonomous Navigation applications since
 - GPS is noisy and doesn't work indoors
 - Gives redundancy, in case GPS fails.

Cambridge Landmarks (Outdoor)



→ 6-DOF
Coordinates

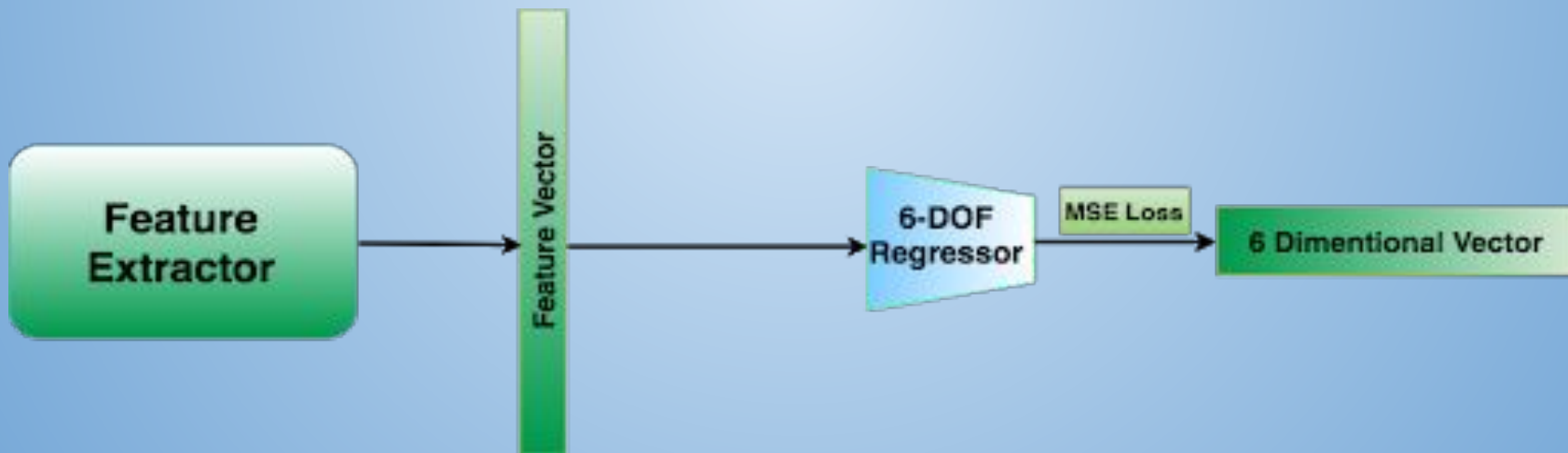
7 Scenes (Indoor)



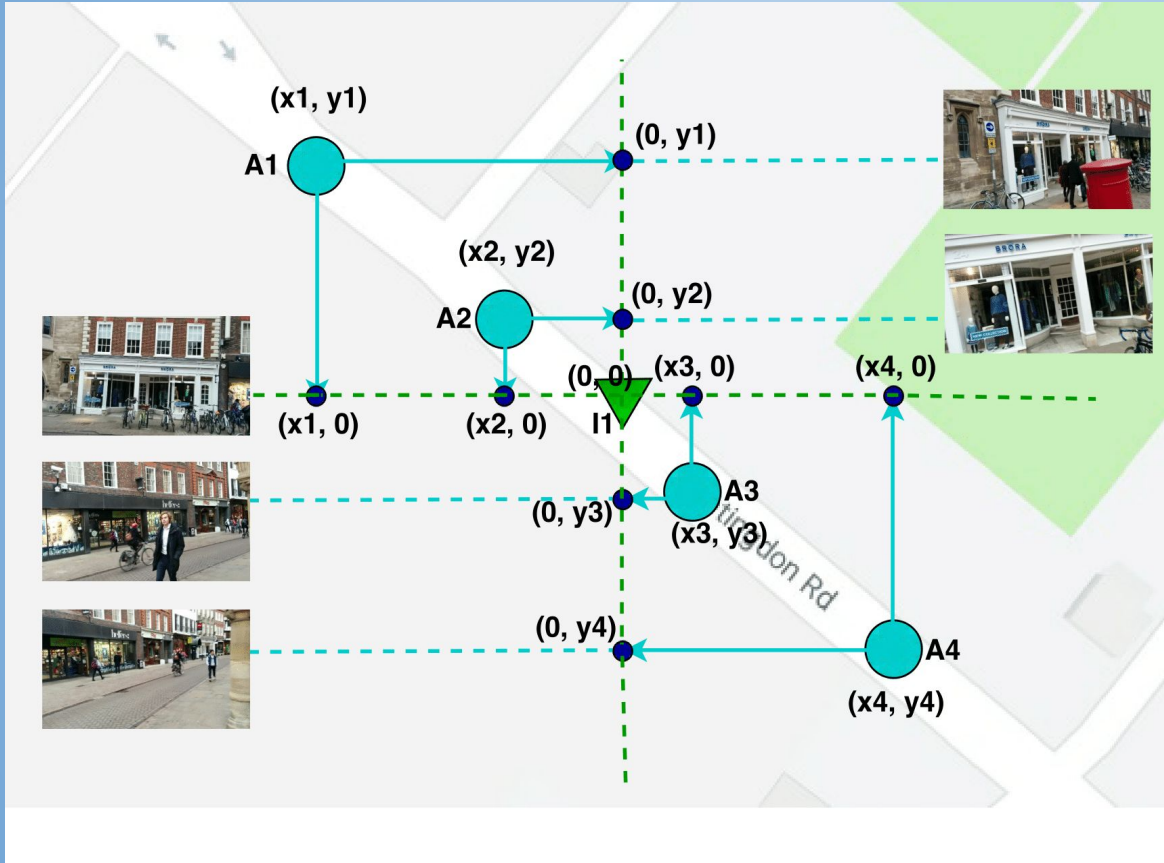
→ 6-DOF
Coordinates

DNN based approaches

- Image Retrieval Based:
 - Extract image features for all images in dataset.
 - Compute nearest neighbour
- PoseNet approach:
 - Use neural network to directly predict 6-DOF.
 - Train DNN by regressing against ground truth 6-DOF in dataset.

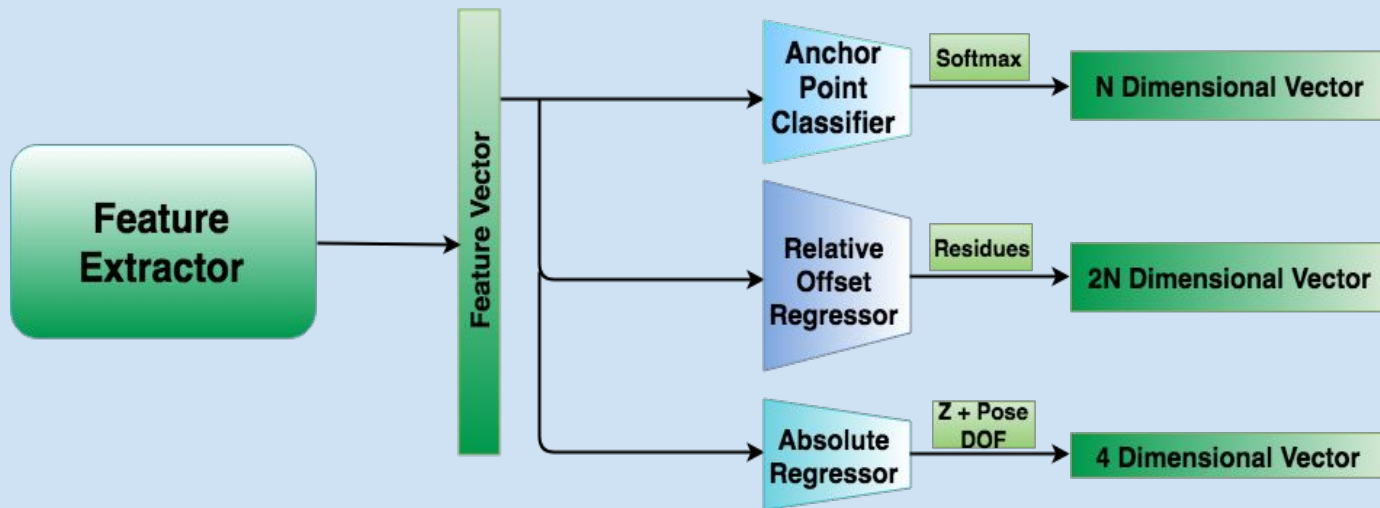


Defining Anchor Points



- Humans typically identify their location relative to other locations or landmarks.
- Inspired from this, we propose an end to end trainable model.
- Define landmarks as anchor points and predict distances relative to them.

Anchor Point based Approach



What is the ground truth for classification?

Is it the nearest anchor point?

Can we discover the most relevant anchor point?

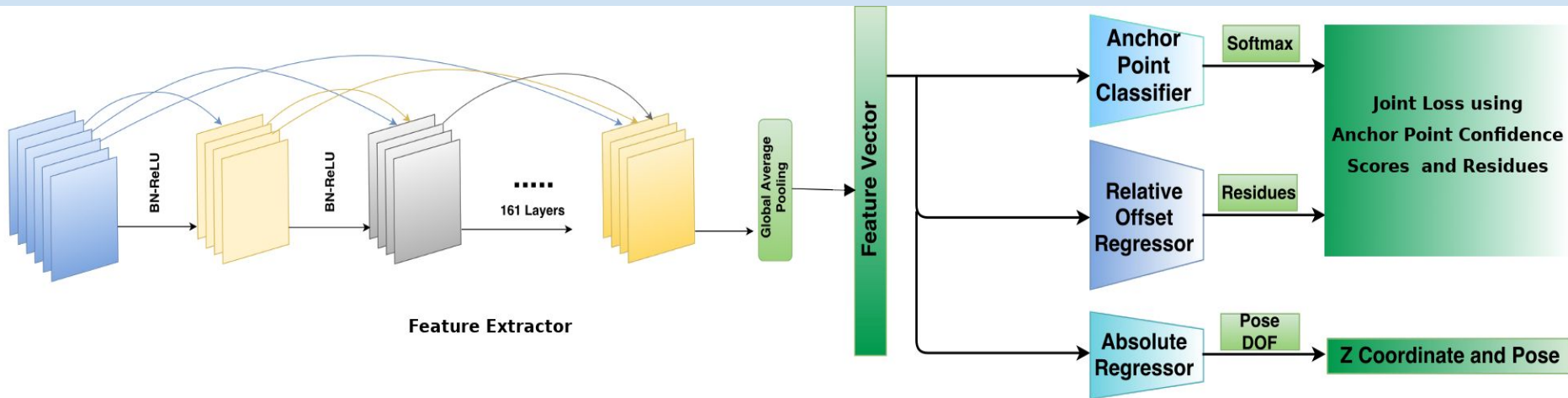
Discovering Relevant Anchor Points

- We have ground truth for all the anchor points.
- We predict the offsets with respect to all of those anchor points.
- Use classification head as confidence score and do a Weighted MSE loss.

<u>Frame Coordinates <X,Y></u>	<u>Anchor points Coordinates <X,Y></u>	<u>Offsets <X,Y></u>
4.25m, 7.51m [Reference Frame]	3.15m, 6.75m [Anchor point 1]	1.10m, 0.76m
	3.60m, 7.00m [Anchor point 2]	0.65m, 0.51m
	4.05m, 7.25m [Anchor point 3]	0.20m, 0.26m
	4.50m, 7.50m [Anchor point 4]	0.25m, 0.01m

Example of offsets to be predicted for 4 Anchor Points

Network Architecture



Loss Function

$$Loss = \alpha_1 H(C_i, \hat{C}_i) + \alpha_2 \sum_i \left[(X_i - \hat{X}_i)^2 + (Y_i - \hat{Y}_i)^2 \right] C_i + \alpha_3 \sum_i \left[(Z_i - \hat{Z}_i) \right]^2 + \left\| P_i - \frac{\hat{P}_i}{\|\hat{P}_i\|} \right\|^2$$

Cross-Entropy
loss
component

Offsets
weighted by
Confidence

Z- coordinate
Regressor

Pose
Regressor










Selected Quantitative Results

<u>Scene</u>	Area or Volume	Posenet Geom. Rep. [11]	Ours (DenseNet) (w/o cross entropy)	Ours (GoogleNet) (w/o cross entropy)
Great Court	8000m ²	6.83m, 3.47°	4.64m, 3.42°	5.89m, 3.53°
King's College	5600m ²	0.88m, 1.04°	0.57m, 0.88°	0.79m, 0.95°
Old Hospital	2000m ²	3.20m, 3.29°	1.21m, 2.55°	2.11m, 3.05°
Shop Facade	875m ²	0.88m, 3.78°	0.52m, 2.27°	0.77m, 3.25°
St. Mary's Church	4800m ²	1.57m, 3.32°	1.04m, 2.69°	1.22m, 3.02°
Street	50000m ²	20.3m, 25.5°	7.86m, 24.2°	11.8m, 24.3°
Chess	6m ²	0.13m, 4.48°	0.06m, 3.89°	0.08m, 4.12°
Fire	2.5m ²	0.27m, 11.3°	0.15m, 10.3°	0.16m, 11.1°
Head	1m ²	0.17m, 13.0°	0.08m, 10.9°	0.09m, 11.2°
Office	7.5m ²	0.19m, 5.55°	0.09m, 5.15°	0.11m, 5.38°
Pumpkin	5m ²	0.26m, 4.75°	0.10m, 2.97°	0.14m, 3.55°
Red Kitchen	18m ²	0.23m, 5.35°	0.08m, 4.68°	0.13m, 5.29°
Stairs	7.5m ²	0.35m, 12.4°	0.10m, 9.26°	0.21m, 11.9°

Selected Quantitative Results

Scene	DenseNet (Feature Extractor)		GoogleNet (Feature Extractor)		MobileNet (Feature Extractor)	
	Performance	FLOPs	Performance	FLOPs	Performance	FLOPs
Kings College	$0.57m, 0.88^\circ$	5998 M	$0.79m, 0.95^\circ$	760 M	$0.67m, 0.94^\circ$	569 M
Shop Facade	$0.52m, 2.27^\circ$		$0.77m, 3.25^\circ$		$0.60m, 2.31^\circ$	

Selected Qualitative Results

Scene	Input Frame	Nearest Anchor Point	Learned Anchor Point
Great Court	 A wide-angle photograph of the Great Court at King's College London, showing a large green lawn in the foreground and the historic Gothic architecture of the buildings in the background under a clear sky.	 A photograph of the Great Court, similar to the input frame, but with a specific region of the lawn and the building facade highlighted as the nearest anchor point.	 A photograph of the Great Court, similar to the input frame, but with a different region of the lawn and the building facade highlighted as the learned anchor point.
King's College	 A photograph of the main entrance of King's College London, featuring a large Gothic archway. A blue van and a white delivery truck are parked in the foreground, and several people are walking.	 A photograph of the main entrance of King's College London, similar to the input frame, but with a specific region of the archway and the building facade highlighted as the nearest anchor point.	 A photograph of the main entrance of King's College London, similar to the input frame, but with a different region of the archway and the building facade highlighted as the learned anchor point.
Old Hospital	 A photograph of a large, multi-story building with a classical facade, identified as the Old Hospital. The building has many windows and is partially obscured by bare trees in the foreground.	 A photograph of the Old Hospital building, similar to the input frame, but with a specific region of the facade highlighted as the nearest anchor point.	 A photograph of the Old Hospital building, similar to the input frame, but with a different region of the facade highlighted as the learned anchor point.

THANK YOU!

VISIT OUR POSTER FOR MORE DETAILS!

Actively looking for PhD opportunities!
sohamsaha.cs@gmail.com