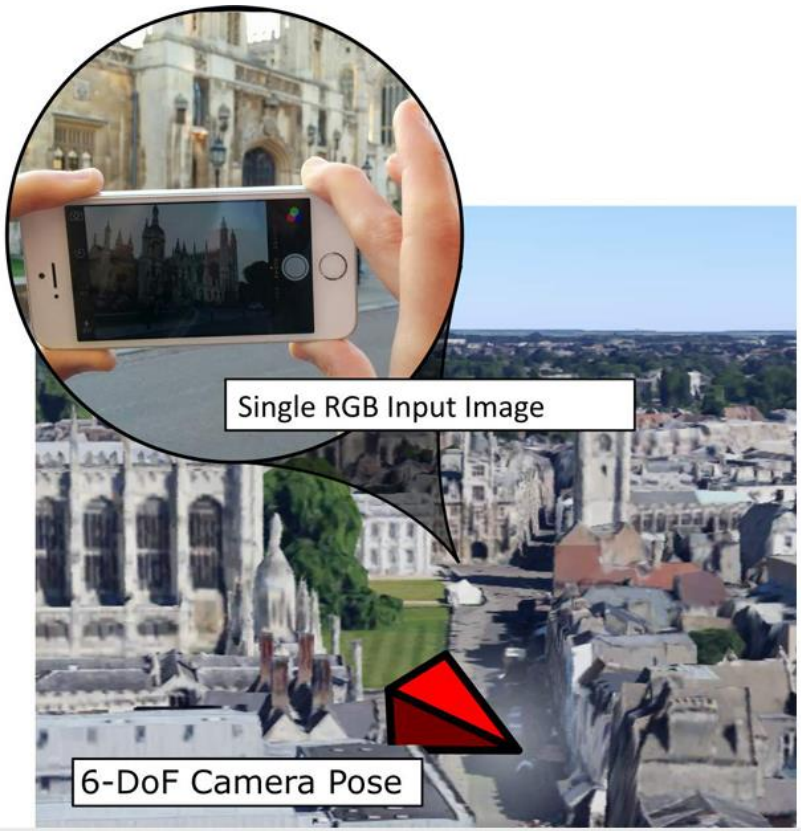
PosNet :
A Convolutional Network for Real-Time 6-DOF Camera Relocalization

Computer Vision & Augmented Reality 연구실
학부연구생 강 준 구

Contents

- Overview
- Dataset
- Architecture
- Conclusion

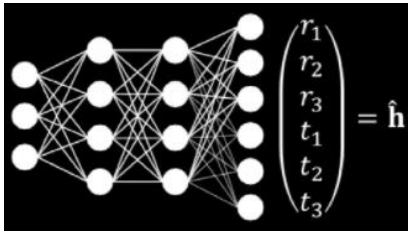


Overview

- Use CNN to regress the camera position and angle

Taking 5ms to run

2m and 6degrees accuracy



Input Image
224x224 RGB image

CNN

6 DOF pose
Six degrees of freedom monocular

Robust to nuisance variables

Reference image

Different scales

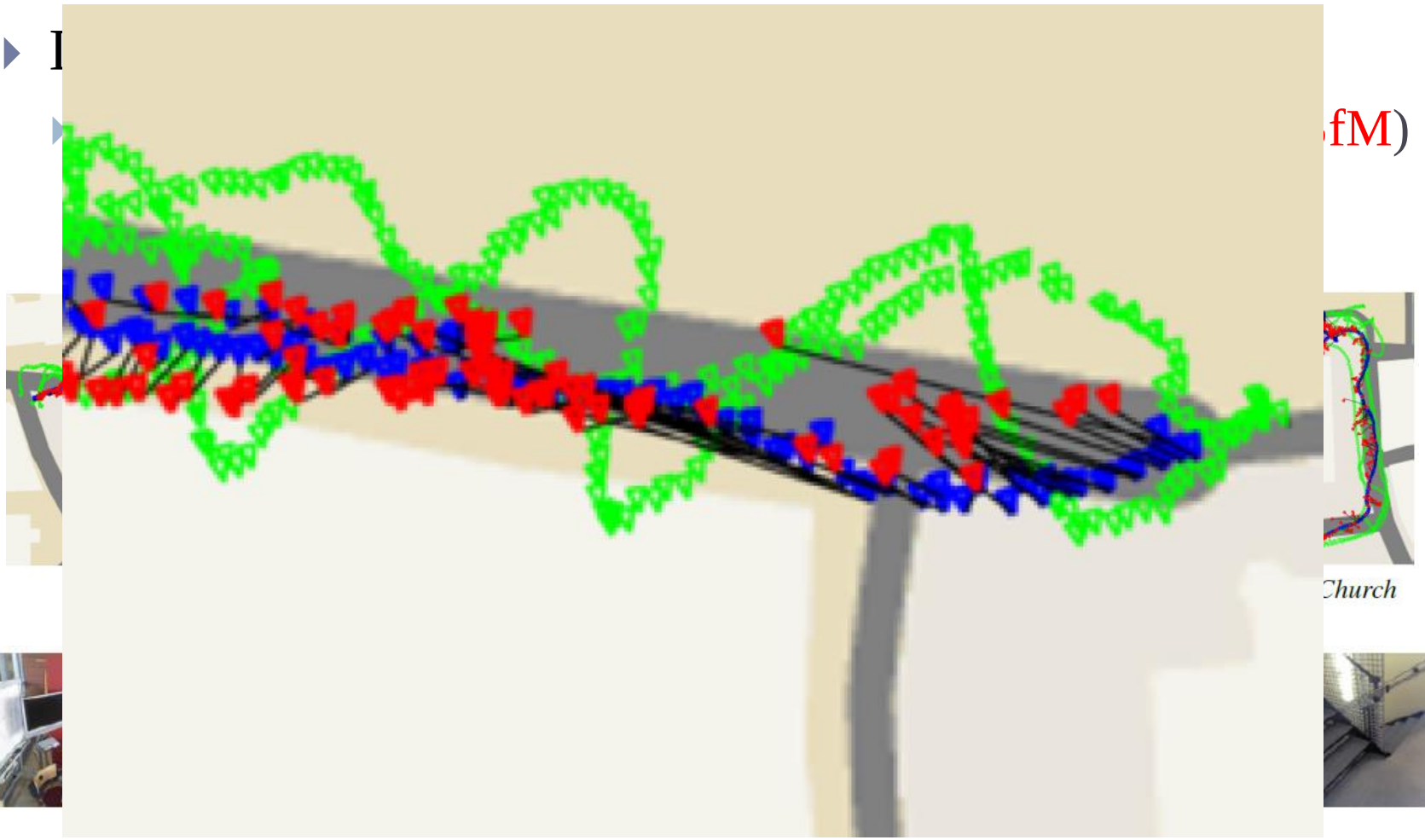
Illumination changes

occlusions

Different seasons

Experimental results

- I
  - fM)



Church

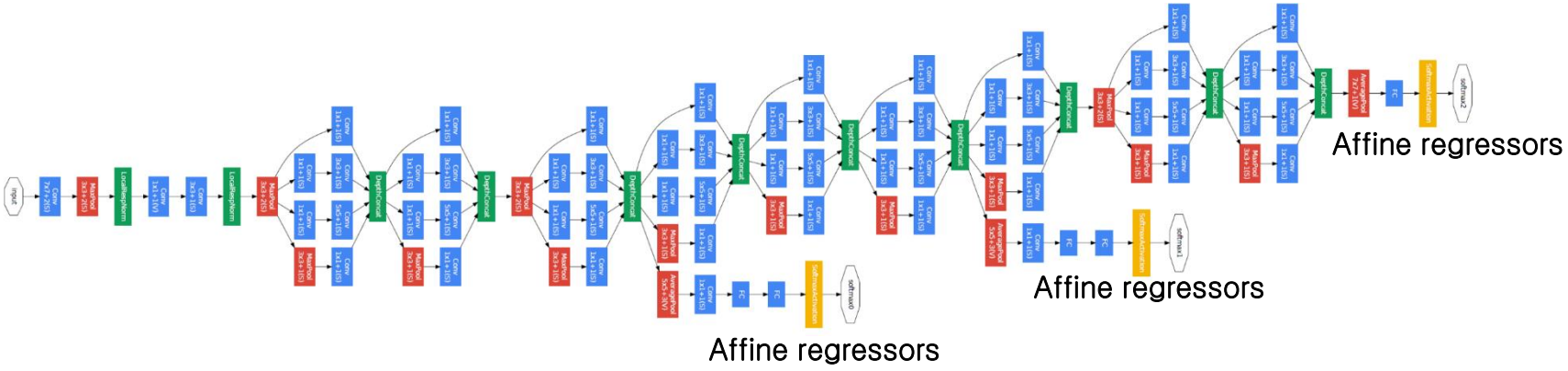
Experimental results

- details and results

Scene	# Frames Train	# Frames Test	Spatial Extent (m)	SCoRe Forest (Uses RGB-D)	Dist. to Conv. Nearest Neighbour	PoseNet	Dense PoseNet
King's College	1220	343	140 x 40m	N/A	3.34m, 5.92°	1.92m, 5.40°	1.66m, 4.86°
Street	3015	2923	500 x 100m	N/A	1.95m, 9.02°	3.67m, 6.50°	2.96m, 6.00°
Old Hospital	895	182	50 x 40m	N/A	5.38m, 9.02°	2.31m, 5.38°	2.62m, 4.90°
Shop Façade	231	103	35 x 25m	N/A	2.10m, 10.4°	1.46m, 8.08°	1.41m, 7.18°
St Mary's Church	1487	530	80 x 60m	N/A	4.48m, 11.3°	2.65m, 8.48°	2.45m, 7.96°
Chess	4000	2000	3 x 2 x 1m	0.03m, 0.66°	0.41m, 11.2°	0.32m, 8.12°	0.32m, 6.60°
Fire	2000	2000	2.5 x 1 x 1m	0.05m, 1.50°	0.54m, 15.5°	0.47m, 14.4°	0.47m, 14.0°
Heads	1000	1000	2 x 0.5 x 1m	0.06m, 5.50°	0.28m, 14.0°	0.29m, 12.0°	0.30m, 12.2°
Office	6000	4000	2.5 x 2 x 1.5m	0.04m, 0.78°	0.49m, 12.0°	0.48m, 7.68°	0.48m, 7.24°
Pumpkin	4000	2000	2.5 x 2 x 1m	0.04m, 0.68°	0.58m, 12.1°	0.47m, 8.42°	0.49m, 8.12°
Red Kitchen	7000	5000	4 x 3 x 1.5m	0.04m, 0.76°	0.58m, 11.3°	0.59m, 8.64°	0.58m, 8.34°
Stairs	2000	1000	2.5 x 2 x 1.5m	0.32m, 1.32°	0.56m, 15.4°	0.47m, 13.8°	0.48m, 13.1°

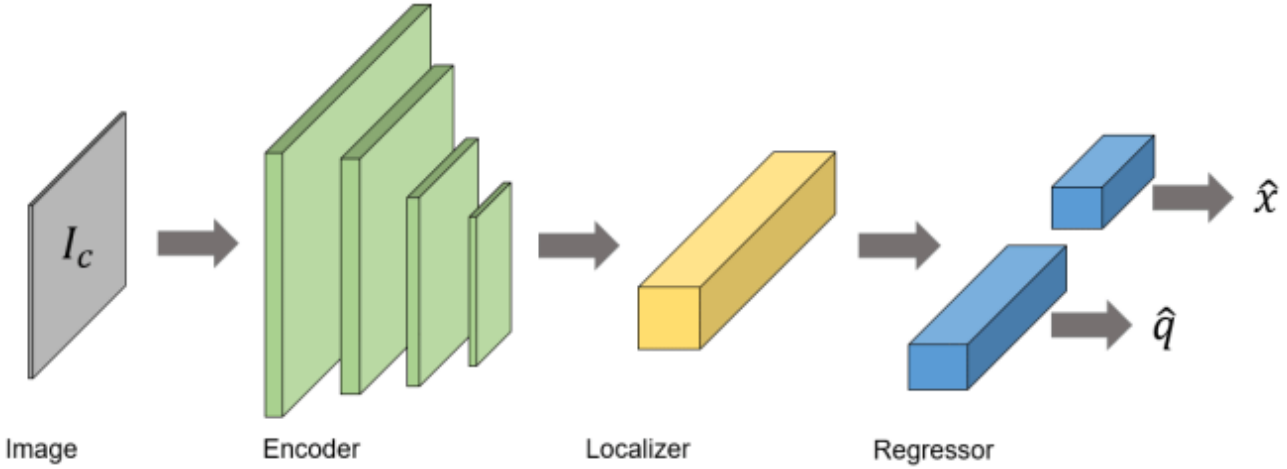
Network

- GoogLeNet
  - The softmax layers were removed
    - -> a pose vector of 7-dimensions
      - Position (3) and orientation (4)



Architecture

- p=[x,q]
  - p : a pose vector relative to an arbitrary global reference frame
    - x : 3D camera position
    - q : Orientation represented by quaternion

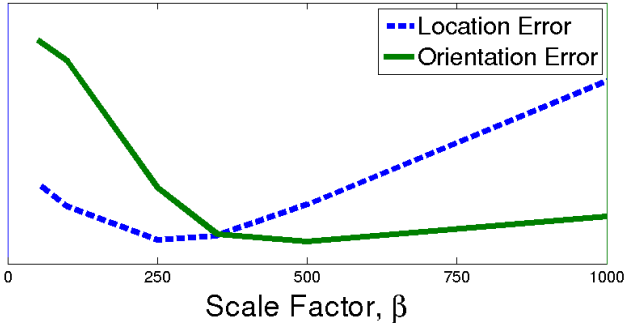


Loss Function

- Learning location and orientation

$$loss(I) = \|\hat{\mathbf{x}} - \mathbf{x}\|_2 + \beta \left\| \hat{\mathbf{q}} - \frac{\mathbf{q}}{\|\mathbf{q}\|} \right\|_2$$

  - Optimal Beta given by ratio between expected error of position and orientation at the end of training (not beginning)

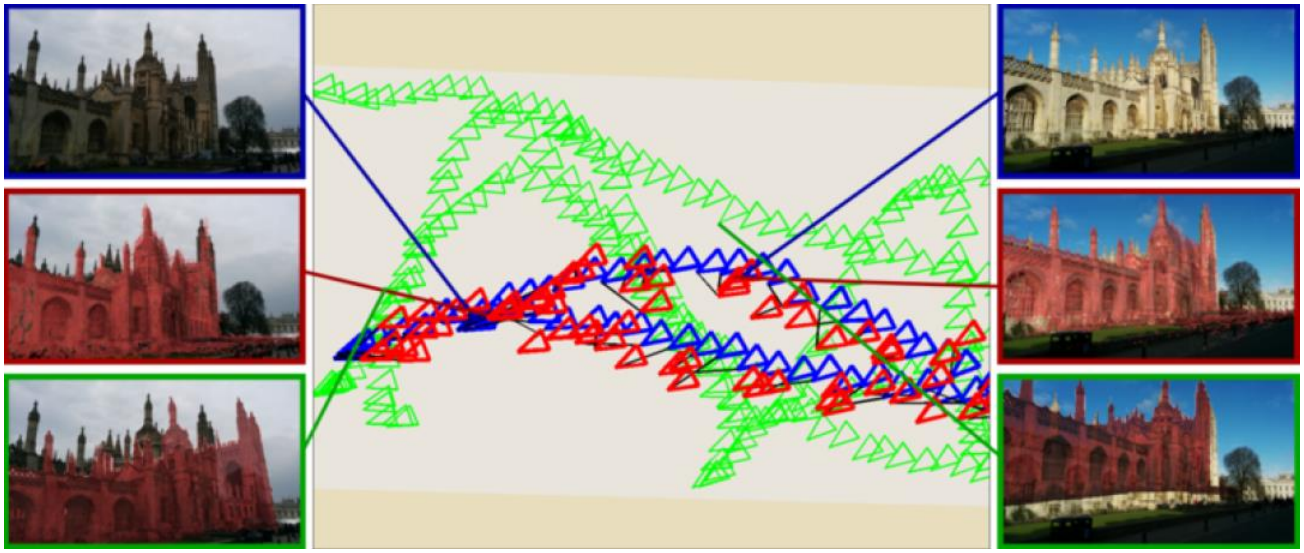


Parameter

- Training
  - SGD optimization
    - Learning rate : 10^-5, Momentum : 0.9
  - Batch size 75
  - Nvidia Titan graphics card training

Review

- Estimate the 3D position and orientation of the camera, given a single monocular image taken from a large previously explored area



Summary

- Conclusion
  - PoseNet is an end-to-end 6DOF pos regression convnet
  - 5ms run-time, 50MB total storage space
  - Robust to lighting, weather, dynamic objects
  - Poor positioning accuracy

Question

- 왜 q에서 |q|분모로 나눠주는가? Unit length?
- 왜 position과 orientation 두개를 학습시키기 어려운가?