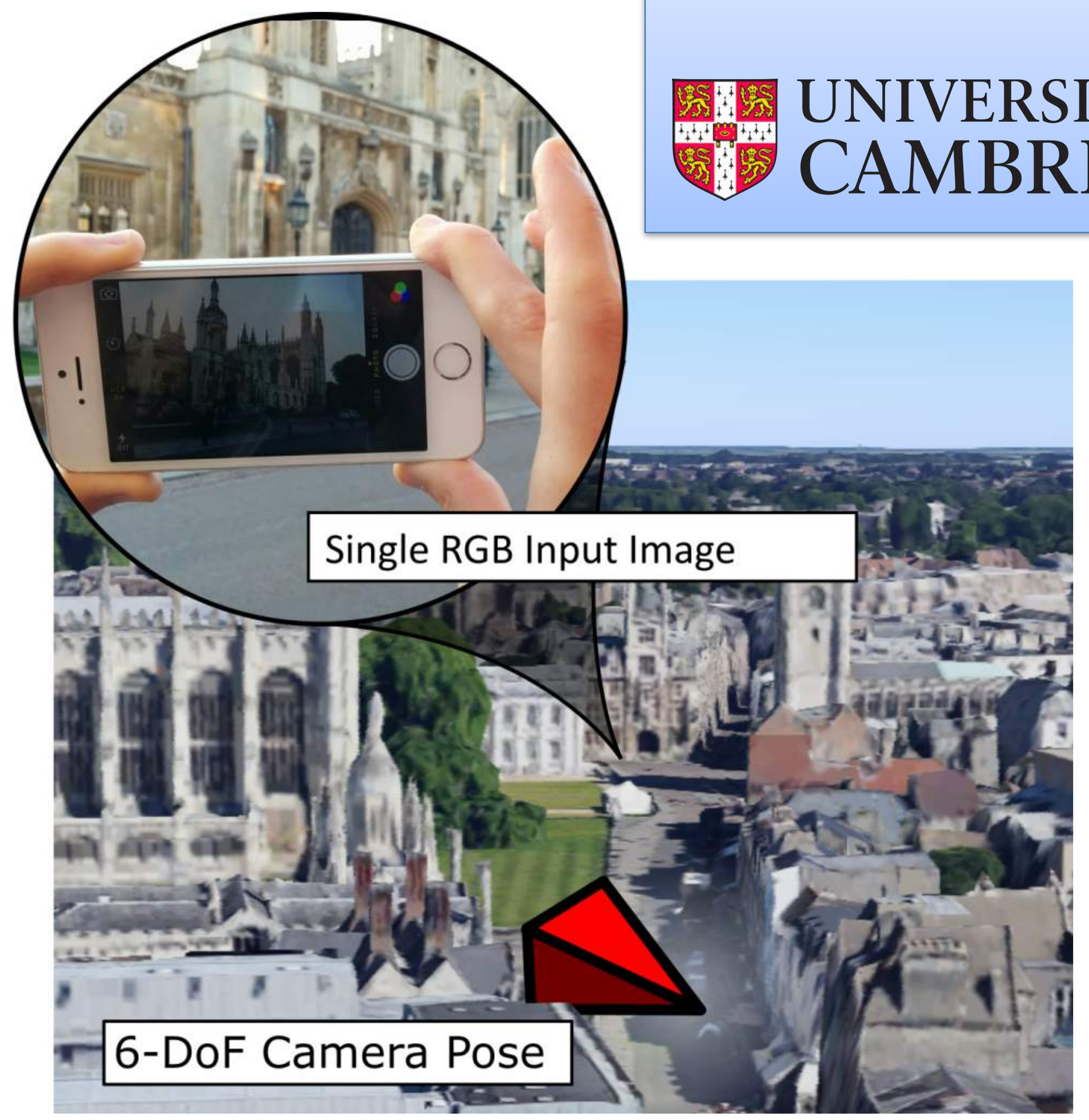
Geometric Loss Functions for Camera Pose Regression with Deep Learning

Alex Kendall and Roberto Cipolla, University of Cambridge

IEEE 2017 Conference on Computer Vision and Pattern Recognition
CVPR July 21-26 2017

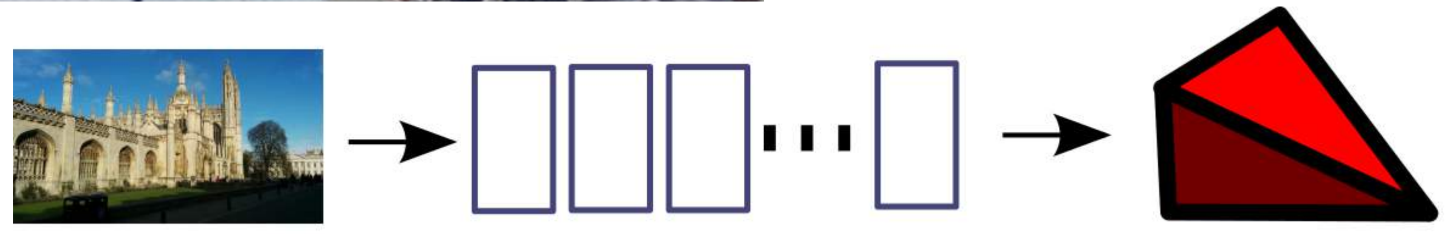
UNIVERSITY OF CAMBRIDGE

Webdemo: http://mi.eng.cam.ac.uk/projects/relocalisation @alexgkendall



The Kidnapped Robot Problem

Is to relocalise within a pre-explored environment.
PoseNet [1] learns a mapping from a single image to 6-DOF camera pose

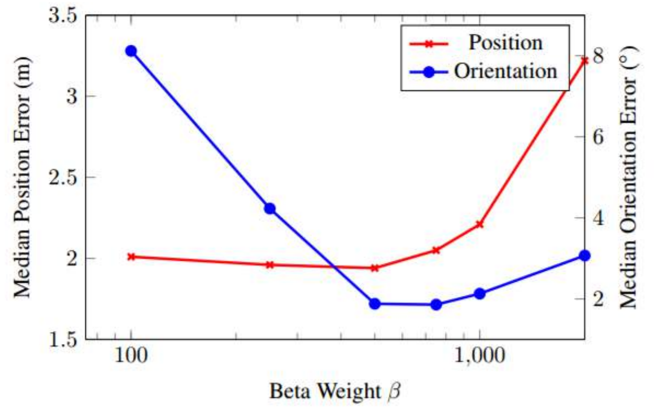


PoseNet

- ✓ Robust to lighting, weather, dynamic objects – learns features based on shape, appearance and global context
- ✓ Fast inference, <2ms per image on Titan GPU
- ✓ Scale not dependent on number of training images
- ✓ Trained with a naïve end-to-end loss function to regress camera position, **x**, and orientation, **q**;

$$\text{loss} = \|x - \hat{x}\|_2 + \beta \left\| q - \frac{\hat{q}}{\|\hat{q}\|} \right\|_2$$

- ✗ Relocalization accuracy of 2m, 5° over scene of 50,000m²… can we do better?
- ✗ How do we weight position, **q**, and orientation, **x**, losses?



This work: Geometric Loss Function

- ➢ Use reprojection function, π, and train on reprojection of 3D geometry in 2D image space
- ➢ Using ideas from bundle adjustment as a differentiable training loss
- ➢ No calibration, we can use arbitrary camera intrinsics

$$\text{loss} = \frac{1}{|\mathcal{G}'|} \sum_{g_i \in \mathcal{G}'} \|\pi(x, q, g_i) - \pi(\hat{x}, \hat{q}, g_i)\|_1$$

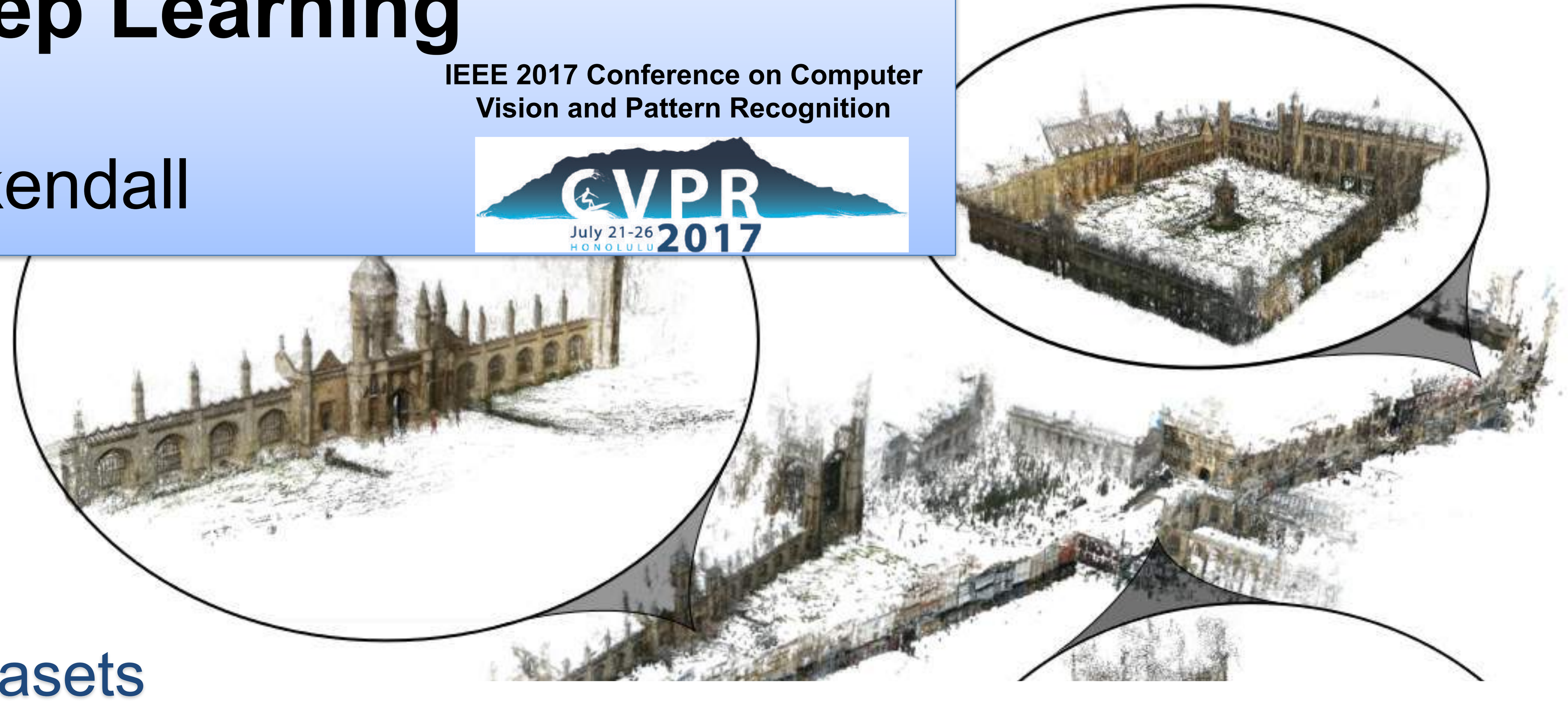
What if we don't have geometry?

- ➢ What can we do if we don't have 3D geometry, e.g. SFM model, RGB-D data
- ➢ We can use task-dependent (homoscedastic) uncertainty to weight position and orientation

$$\text{loss} = \frac{\|x - \hat{x}\|_2}{\sigma_x^2} + \log \sigma_x^2 + \frac{\|q - \hat{q}/\|\hat{q}\|\|_2}{\sigma_q^2} + \log \sigma_q^2$$

Performance

	Cambridge Landmarks, King's College			Dubrovnik 6K		
	Median Error		Accuracy	Median Error		Accuracy
Loss function	x[m]	q[°]	< 2m,5°	x[m]	q[°]	< 2m,5°
Linear sum, β = 500 [1]	1.52	1.19	65%	13.1	4.68	30.1%
Learn weighting with task uncertainty	0.99	1.06	85.3%	9.88	4.73	41.7%
Reprojection loss	*does not converge*					
Learn weighting pretrain + Reprojection loss	0.88	1.04	90.3%	7.90	4.40	48.6%
SIFT + SfM Geometry [4]	0.42	0.55	-	1.1	-	-



Datasets

- ➢ Cambridge landmarks, 100 x 500m street scenes, Kendall et al.

- ➢ Seven Scenes, 4 x 3m indoor scenes, Shotton et al.

- ➢ Dubrovnik Dataset, 1500 x 1500 m small town, Li et al.

Future Work:

- ➢ City-scale metric localisation
- ➢ Fine grained localisation – achieve accuracy which enables augmented reality
- ➢ Temporal localisation and end-to-end learning for SLAM

References:

[1] Alex Kendall, Matthew Grimes and Roberto Cipolla. PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization. ICCV, 2015.
[2] Alex Kendall and Roberto Cipolla. Modelling Uncertainty in Deep Learning for Camera Relocalization ICRA, 2016.
[3] Alex Kendall and Roberto Cipolla. Geometric loss functions for camera pose regression with deep learning. CVPR, 2017.
[4] Torsten Sattler, Bastian Leibe, and Leif Kobbelt. Efficient & effective prioritized matching for large-scale image-based localization. PAMI, 2016.

CVPR Tutorial

Large-Scale Visual Place Recognition and Image-Based Localization
Wednesday, July 26th, 2017 - morning (half-day)