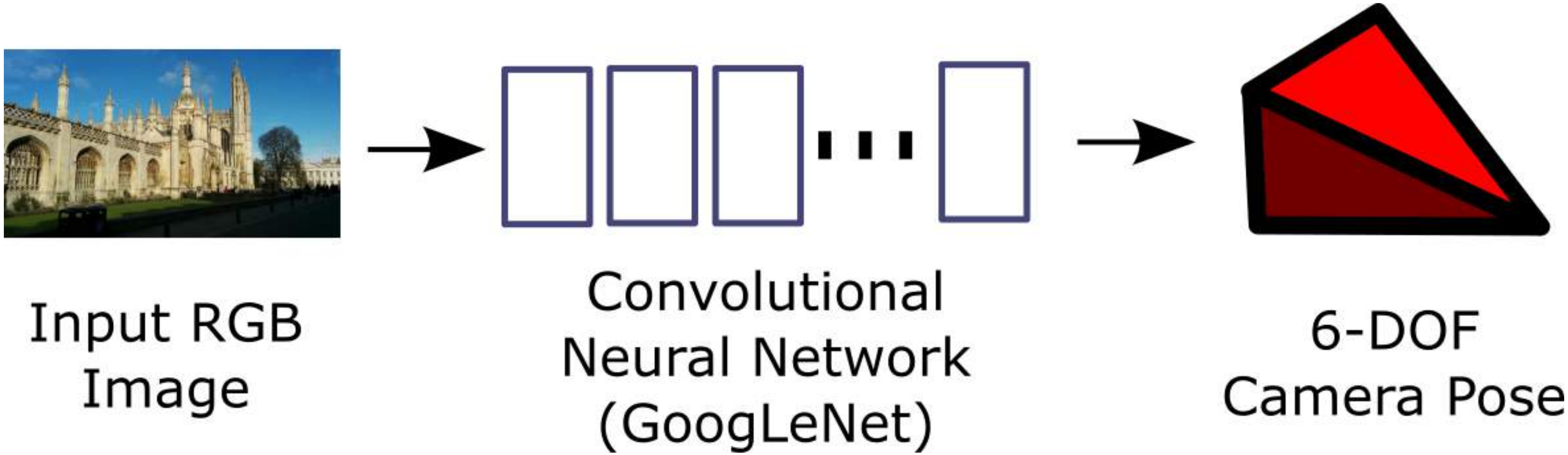
UNIVERSITY OF CAMBRIDGE

Geometric loss functions for camera pose regression with deep learning

Alex Kendall and Roberto Cipolla, University of Cambridge

@alexgkendall mi.eng.cam.ac.uk/projects/relocalisation/



Trained with a naïve end-to-end loss function to

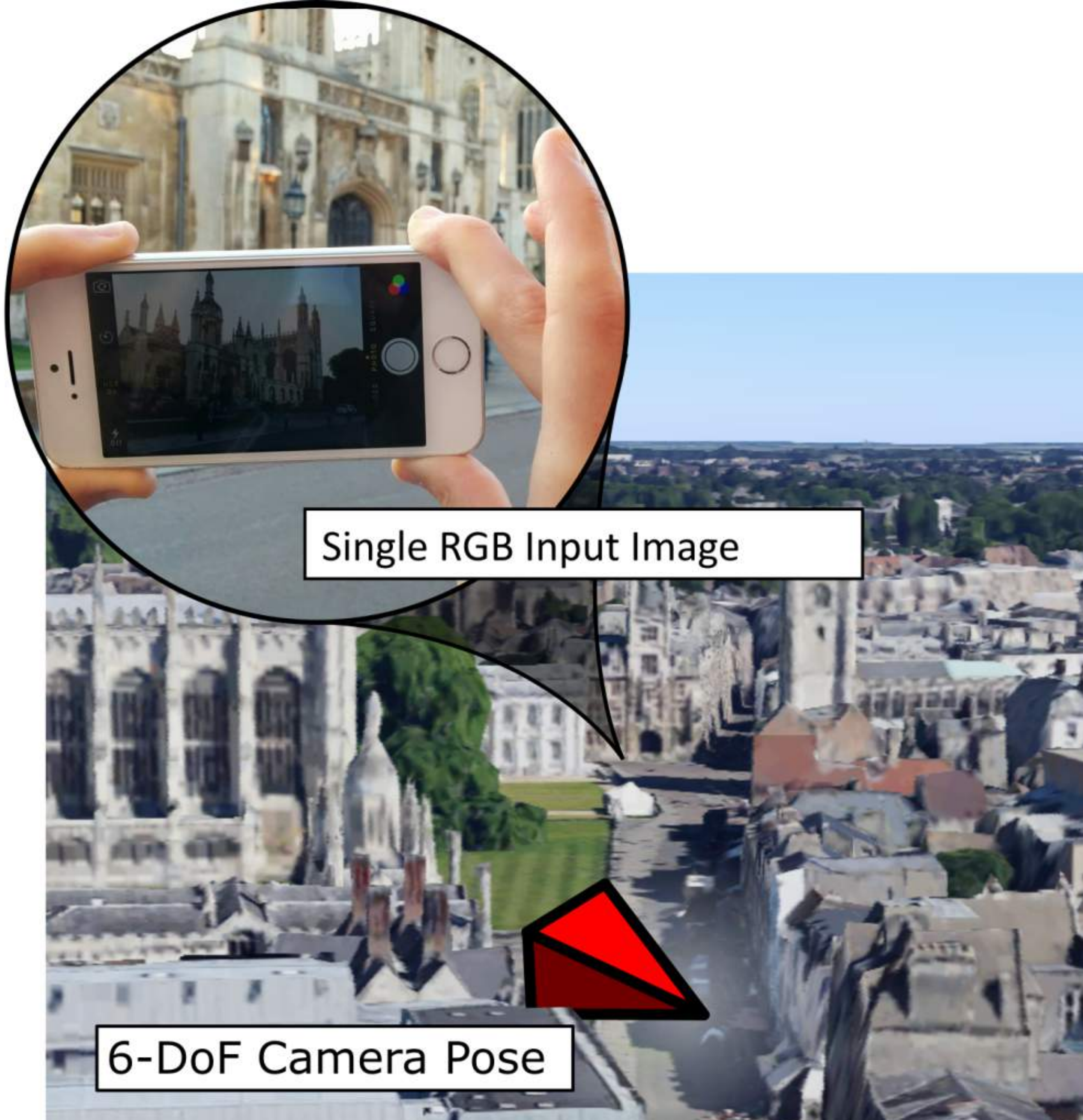
regress camera position, **x**, and orientation, **q**

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

Alex Kendall, Matthew Grimes and Roberto Cipolla. **PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization.** ICCV, 2015.

PoseNet: Learning to Localize

- Robust to lighting, weather, dynamic objects
- Fast inference, <2ms per image on Titan GPU
- Scale not dependent on number of training images

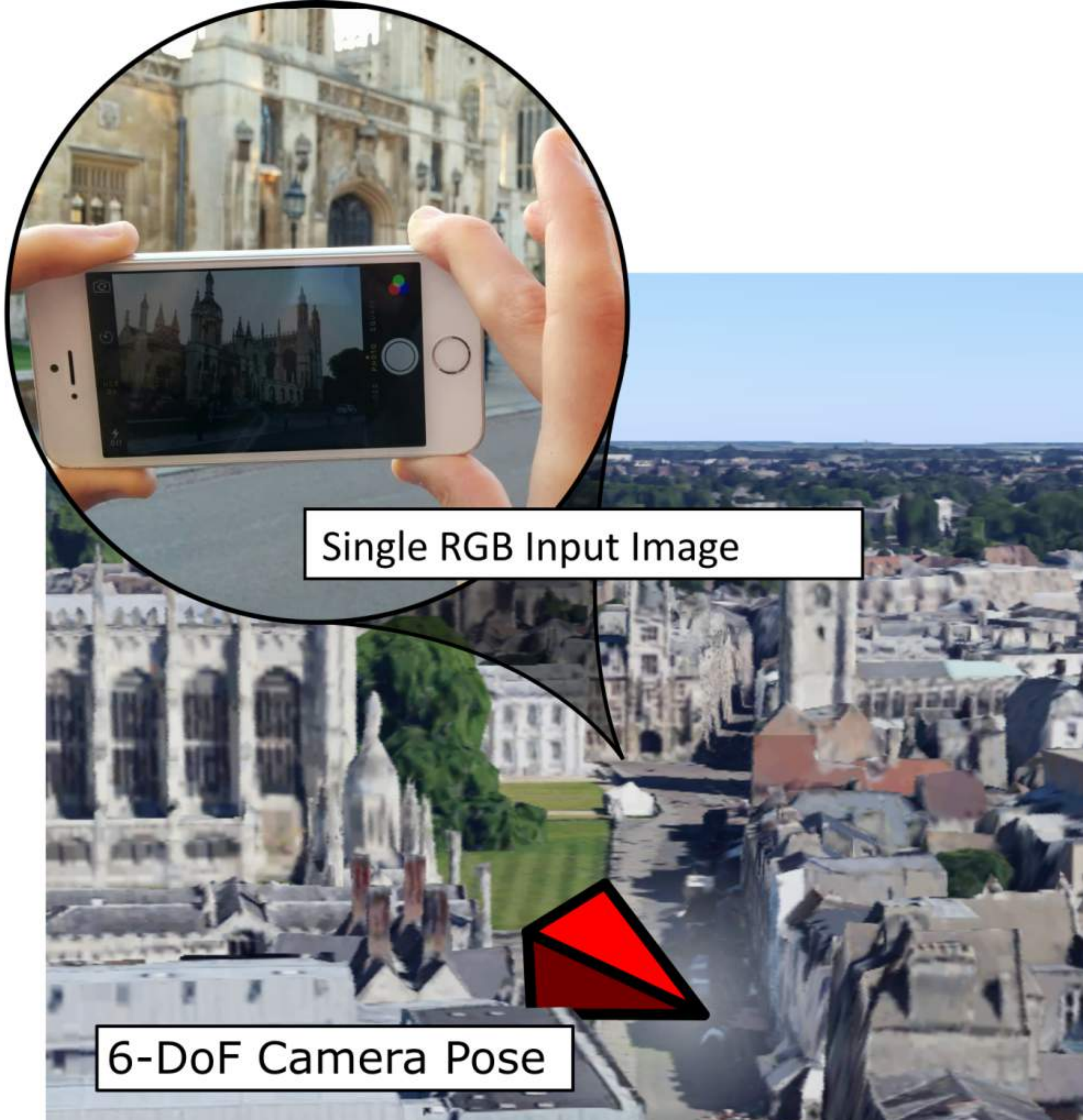


Alex Kendall, Matthew Grimes and Roberto Cipolla. **PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization.** ICCV, 2015.

Problems?

1. How do we weight position, q, and orientation, x, losses?
2. Relocalization accuracy of 2m, 5° over scene of 50,000m^2… can we do better?!
3. Coarse accuracy is not sufficient for fine grained localisation tasks e.g. augmented reality

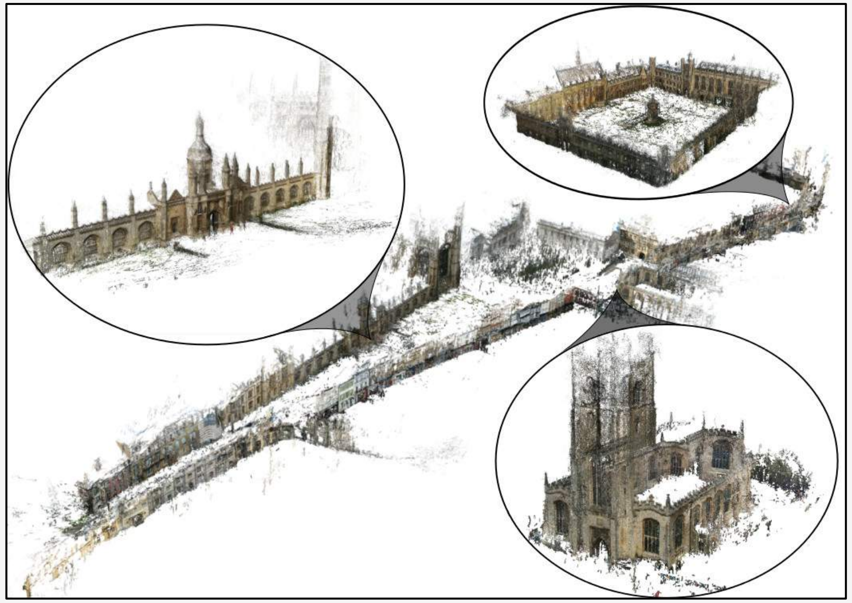
Alex Kendall, Matthew Grimes and Roberto Cipolla. **PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization.** ICCV, 2015.



This work: Learning camera pose, *with geometry*

Train with reprojection loss of 3-D geometry using predicted and ground truth camera poses.

$$loss(I) = \frac{1}{|\mathcal{G}'|} \sum_{g_i \in \mathcal{G}'} \left\| \pi(\mathbf{q}, \mathbf{x}, \mathbf{g_i}) - \pi(\hat{\mathbf{q}}, \hat{\mathbf{x}}, \mathbf{g_i}) \right\|_\gamma$$



Where π is the projection function of 3-D point g_i

Automatically learns a weighting between position and orientation!

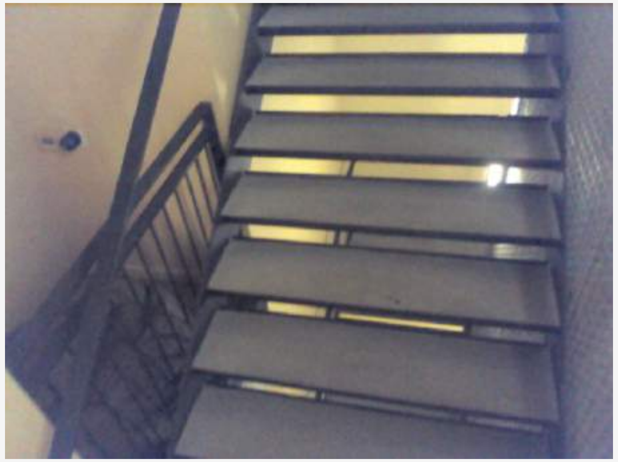
Alex Kendall and Roberto Cipolla. **Geometric loss functions for camera pose regression with deep learning.** CVPR, 2017.

Datasets – Cambridge Landmarks – Outdoor Localization

- 8,000 images from 6 scenes up to 100 x 500m

Alex Kendall, Matthew Grimes and Roberto Cipolla. **PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization.** ICCV, 2015.

Datasets – Seven Scenes – Indoor Localization




- 17,000 images across 7 small indoor scenes.

Jamie Shotton et al. Scene coordinate regression forests for camera relocalization in RGB-D images. CVPR 2013

Datasets – Dubrovnik – Large Scale Localization

- 6000 images across 1500 x 1500 m in Dubrovnik, Croatia.
- Varying weather, season, camera type

Li, Yunpeng, Noah Snavely, and Daniel P. Huttenlocher. "Location recognition using prioritized feature matching." ECCV, 2010.

Geometry Improves Performance

Dataset	Environment	PoseNet	PoseNet with Geometry
Cambridge Landmarks	Street Scenes	2.0m, 6.2°	**1.6m, 2.9°**
7 Scenes	Indoor Rooms	0.45m, 10.0°	**0.23m, 8.1°**
Dubrovnik	Town	13.1m, 4.7°	**7.9m, 4.4°**

Alex Kendall and Roberto Cipolla. **Geometric loss functions for camera pose regression with deep learning.** CVPR, 2017.

Future Work & What's Next?

- PoseNet is much faster and requires smaller images than traditional methods

Dataset	PoseNet with Geometry [1]	Active Search (SIFT + Geometry) [2]
King's College	0.88m, 1.04°	**0.42m, 0.55°**
Resolution	**256 x 256**	1920 × 1080 MP
Inference Time	**2 ms**	78 ms

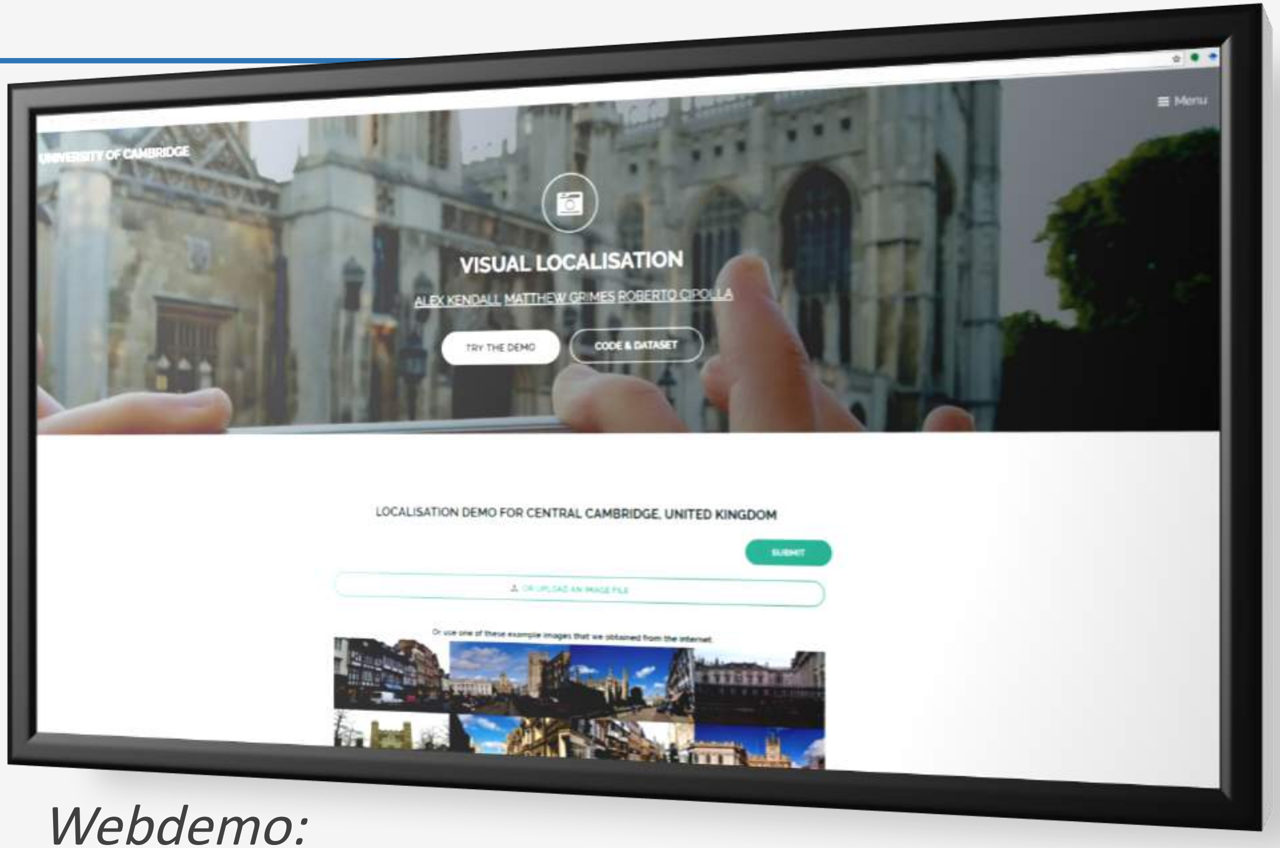
- Can we improve model towards city scale localisation with deep learning
- Improve fine grained accuracy for accurate registration

[1]. Alex Kendall and Roberto Cipolla. **Geometric loss functions for camera pose regression with deep learning.** CVPR, 2017.

[2]. T. Sattler, B. Leibe, and L. Kobbelt. Efficient & effective prioritized matching for large-scale image-based localization. PAMI, 2016.

More to discuss at our poster!

1. What to do if geometry isn't available?
2. Modelling uncertainty
3. Learning rotation representation



@alexgkendall

alexgkendall.com

Webdemo:
mi.eng.cam.ac.uk/projects/relocalisation/

More at CVPR Tutorial on Large Scale Localisation, July 26th (morning)

Alex Kendall and Roberto Cipolla. **Geometric loss functions for camera pose regression with deep learning.** CVPR, 2017.