Learning How to Learn Where You Are: Meta-Learning for Few-Shot Camera Localization

Dominik Kloepfer, João Henriques, Dylan Campbell

Visual Geometry Group, University of Oxford

Problem Setting

- Show the algorithm RGB-D images of a scene together with ground-truth camera poses
- Predict the camera pose from which an unseen RGB image of the same scene was taken

Background

Camera Localization

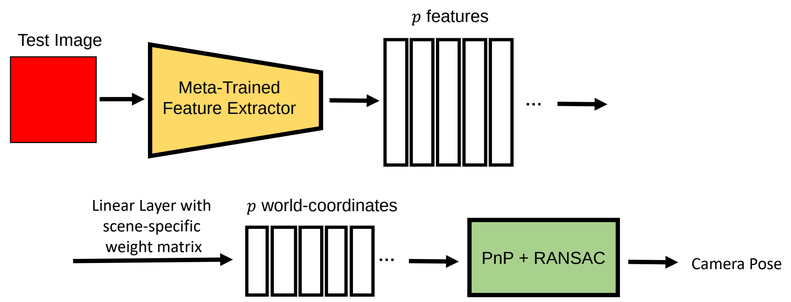
- Previous methods had different drawbacks
- *Absolute Pose Regression* models require large training sets and need to be retrained for each new scene
- *Structure-based* methods are fast and accurate, but usually require a lot of training data or detailed 3D-scans of a scene

Meta-Learning (“Learning how to learn”)

- Two nested training loops
- A *base-learner* works at the level of individual episodes
- A *meta-learner* learns from a collection of such episodes and is trained to improve the performance of the base-learner.

Method

- A combination of DSAC++ [1] and meta-learning to reduce the number of images *n* required to learn a scene
- A meta-trained CNN followed by a scene-specific linear layer (trained using Ridge Regression, as in [2]) predicts the 3D world-coordinates of *p* pixels
- These predictions are used by a Perspective-n-Point algorithm in a RANSAC loop to predict the camera pose



Experiments

- Using the 7Scenes [3] dataset of seven indoor scenes

Results

- We show the algorithm 128 images of a scene to train it (previous methods use several thousand images to learn a scene)
- After meta-training, average error in predicting 3D-world coordinates on new scenes was around 20cm, with the median error at 5-10cm
- Camera pose predictions however failed completely
- This might be due to a bug in the PnP algorithm or the RANSAC implementation
- It is also possible that these algorithms amplify errors so that the coordinate-prediction was not accurate enough

References

[1] Eric Brachmann and Carsten Rother. Learning less is more - 6d camera localization via 3d surface regression. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4654–4662, 2018.

[2] Luca Bertinetto, Joao F. Henriques, Philip Torr, and Andrea Vedaldi. Meta-learning with differentiable closed-form solvers. In *International Conference on Learning Representations*, 2019.

[3] Jamie Shotton, Ben Glocker, Christopher Zach, Shahram Izadi, Antonio Criminisi, and Andrew Fitzgibbon. Scene coordinate regression forests for camera relocalization in rgb-d images. In *Proc. Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2013