# Semantic Clustering of Image Retrieval Databases used for Visual Localization

Henry Hölzemann, Torsten Fiolka

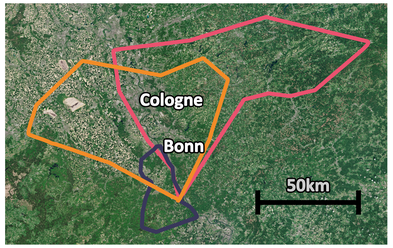
Fraunhofer FKIE

WACV 2025 TUCSON, ARIZONA • FEB 28 - MAR 4

## Background and Setting

**Problem**
- Retrieval-based Visual Localization in large-scale GNSS-denied areas suffers from vast search spaces
- Full search not feasible on small UAS

**Key ideas**
- Partition search space semantically, each group only contains images with similar semantics
- Restrict search to semantically similar groups




**4 Seasons (4S) Dataset**
- Large-scale dataset for Visual Localization and SLAM
- Collected from microlight
- Trajectories repeatedly flown across multiple seasons

**Corine Land Cover (CLC)**
- Pan-European land cover inventory providing geospatial annotations for training semantic segmentation models

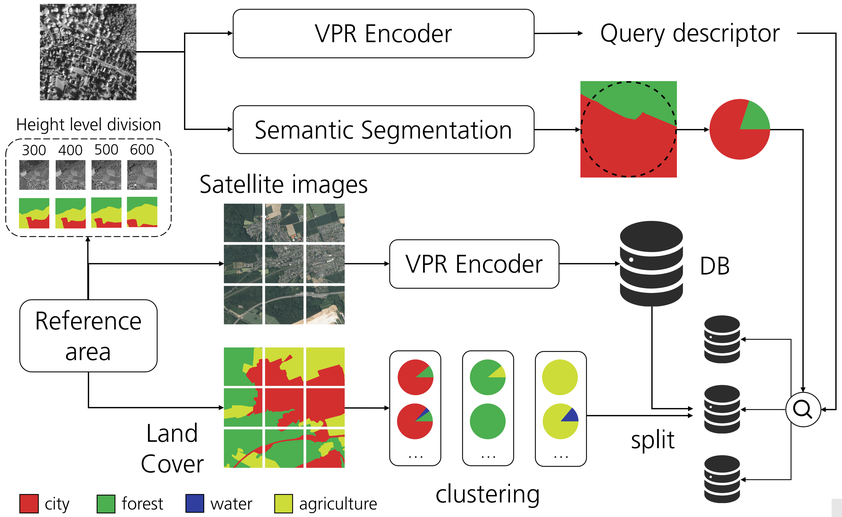
RTK-GNSS groundtruth information → Geospatial bounding box → CLC →



Semantic training data

**Semantic 4S and 4S for VPR**
- 16.8k annotated segmentation masks for training, 4.4k for val.
- 11.2k training queries, 2.2k val queries for VPR
- Grid structure for satellite references used for positive references and distractor images in training and validation

## Objective and Approach




**Semantic Segmentation of grayscale aerial images**
- DeepLabv3+ pretrained on ImageNet and MS COCO in grayscale
- Weighted CE-Loss and Median Frequency Balancing

**Semantic Clustering**
- Cluster VPR reference database based on land cover distributions
- Maximize reduction of search space by utilizing balanced clustering

$$r = 1 - \max_{i \in [1:k]} \frac{|C_i|}{\sum_j |C_j|}$$

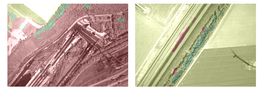
- Soft-assignments to multiple clusters increase robustness in border regions

$$\gamma_i = 1 - \frac{\|D_q - m_i\|_2}{\sum_j \|D_q - m_j\|_2}$$

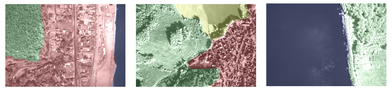
**Retrieval-based VPR**
- NetVLAD as VPR encoder to create descriptors for clustered DBs

## Experiments and Results

**Semantic Segmentation**
- Pixel acc. of 95.2%, mIoU of 91.7% on Semantic 4S val. split



Poor qualitative results

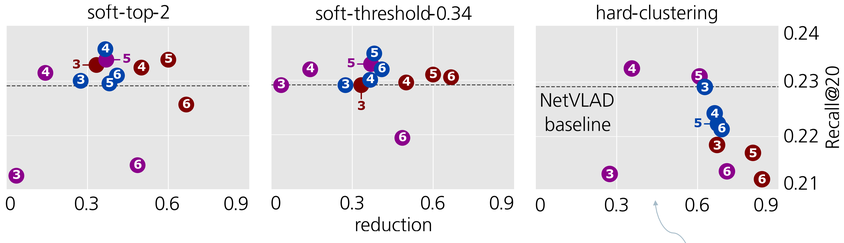


Adequate qualitative results

**Vanilla retrieval-based VPR**
- Results on 4S VPR val. split with 2k queries and 10k distractors

| | *With height division* | | | | *Without height division* | | | |
|---|---|---|---|---|---|---|---|---|
| Backbone | R@1 | R@5 | R@10 | R@20 | R@1 | R@5 | R@10 | R@20 |
| ResNet18 | 4.1% | 10.2% | 13.7% | 21.0% | 3.9% | 10.2% | 11.7% | 14.0% |
| ResNet50 | 4.7% | 14.0% | 18.0% | 22.9% | 5.0% | 11.1% | 14.7% | 19.1% |

**Hybrid VPR**
- Semantic Clustering results for Bal. k-means, k-means and GMMs




- Image retrieval performance evaluation

**Vanilla retrieval time:** 129ms, **retrieval times for clustered DBs:**

| | *Balanced k-means* | | *k-means* | | *GMMs* | |
|---|---|---|---|---|---|---|
| k | reduct. | retr. time | reduct. | retr. time | reduct. | retr. time |
| 3 | 67% | 43ms | 62% | 41ms | 27% | 59ms |
| 4 | 75% | 32ms | 66% | 34ms | 35% | 35ms |
| 5 | 80% | 26ms | 67% | 31ms | 60% | 29ms |
| 6 | 83% | 22ms | 68% | 27ms | 70% | 24ms |

Changes in recall@20 when using annotated segmentation for cluster assignment:

