EVALUATION CRITERIA FOR INSIDE-OUT INDOOR POSITIONING SYSTEMS BASED ON MACHINE LEARNING

Christoffer Löffler[1], Sascha Riechel[1], Janina Fischer[1], and Christopher Mutschler[1 2]

1 Fraunhofer Institute for Integrated Circuits IIS, Precise Localization and Analytics Dept., Machine Learning and Data Analytics Group, Erlangen, Germany
2 University of Erlangen-Nuremberg, Computer Science Department, Machine Learning and Data Analytics Group, Erlangen, Germany

Fraunhofer

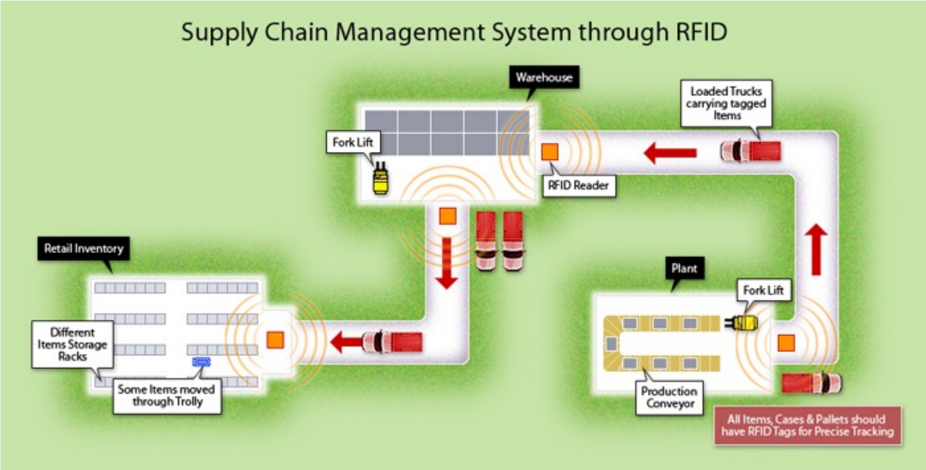
MACHINE LEARNING & DATA ANALYTICS

IPIN 2018
NINTH INTERNATIONAL CONFERENCE ON INDOOR POSITIONING AND INDOOR NAVIGATION
SEPTEMBER 24-27, 2018, NANTES, FRANCE

Motivation – Why do we do this?

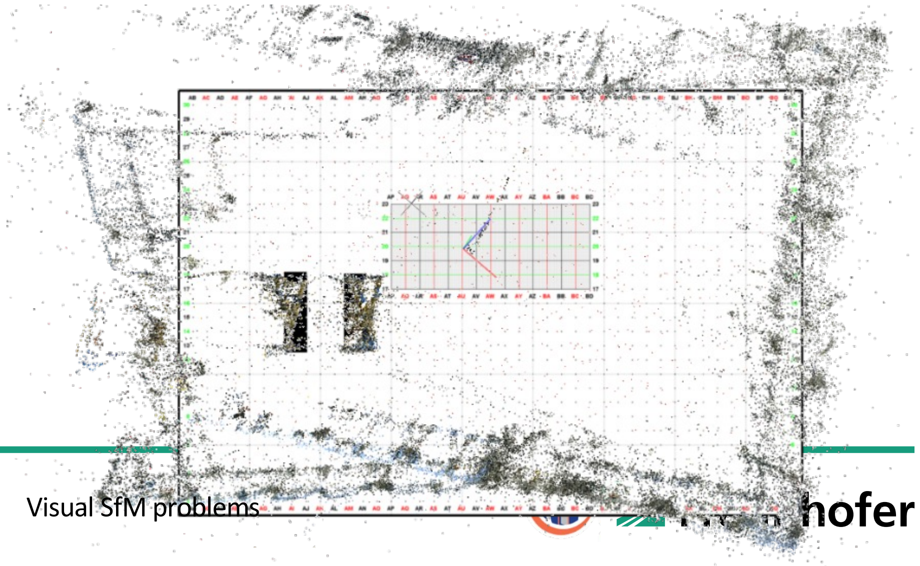
Self-positioning systems in a logistics environment

- **Real-time tracking** in logistics environment (e.g. indoor warehouses) is often used for tracing goods and optimizing the logistics processes
  - **Radio-based or camera-based outside-in positioning** has a high setup effort and is impractical in some scenarios (e.g. occlusion)
  - **Camera-based inside-out self-positioning** mostly utilizes hand-crafted markers or natural features such as edges
    - ➢ Both approaches require **expensive feature detection**
    - ➢ **Changes in the environment** result in poor performance
    - ➢ Structure from Motion (SfM) sometimes fails completely
  - Some camera-based systems **rely on assistance** from additional knowledge (e.g. 3D Models) or sensors (e.g. RGB-D cameras)
    - ➢ **Costly** and **limited usability** (e.g. distance limits)



Hammad Tariq [10]

Mautz et al. (2011) [1]

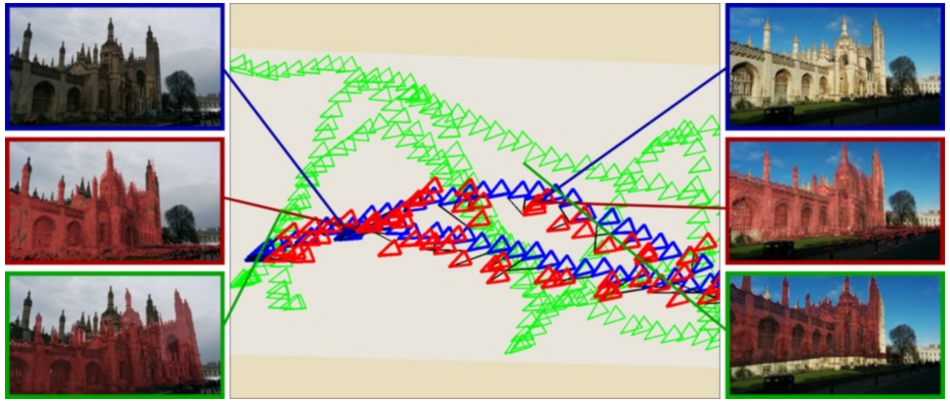


Visual SfM problems

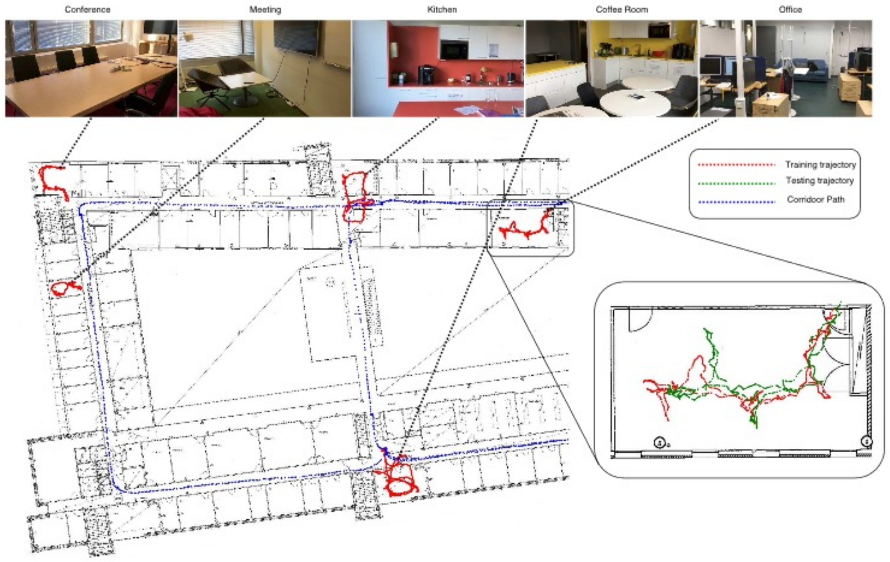
Motivation – Why do we do this?

Application of self-positioning systems in a logistics environment

- Recent advances in machine learning, such as regression forests and **deep convolutional neural networks** (CNNs) become an alternative
  - Example: **PoseNet** by Kendall et al. (2015) [2]
  - Deep Learning architectures learn relevant features from the images and regress the according position
- **Advantages** of such systems
  - No markers or additional sensors needed
  - Robustness to previously unseen or blurry images
- ...**however**, these are often tested outside or in small environments
  - Questions from an indoor positioning point of view remain open (e.g. lighting conditions, varying environment scales, etc.)



PoseNet by Kendall et al. (2015)

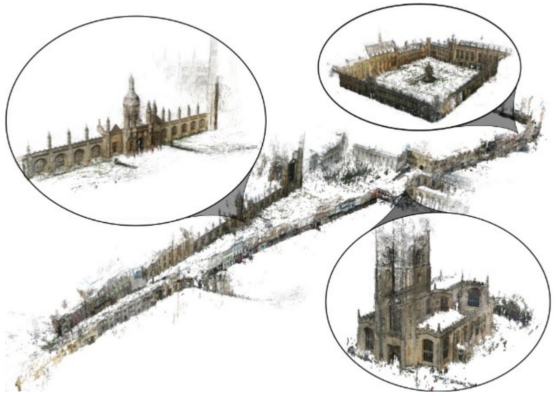


Pairwise Relative Poses, by Laskar et al. (2017)

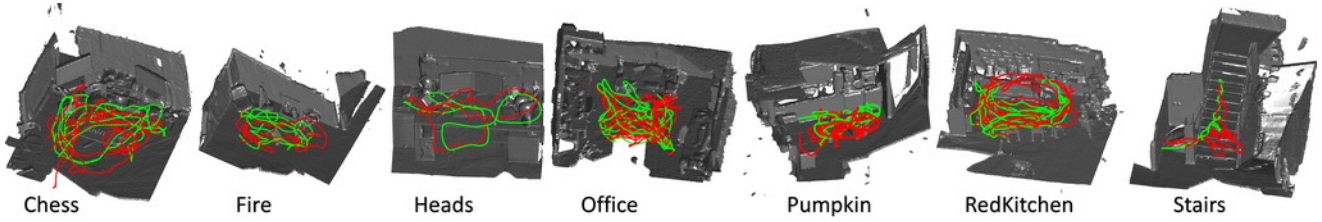
Fraunhofer

Related Work

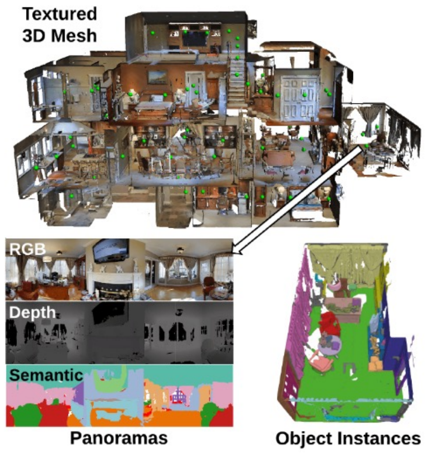
Camera-based Positioning Evaluation Datasets



- Outdoor Datasets
  - Cambridge Landmark [2]
- Small Indoor Datasets
  - 7 Scenes [4]
  - University [3]



- Large Indoor Datasets
  - Baidu [5]
  - Matterport3D [6]
  - Wijams [7]
  - InLoc [8]
  - TU Munich Large Scale-Indoor dataset [9]



Fraunhofer

Related Work
Camera-based Positioning Evaluation Datasets

- Outdoor Datasets
  - Cambridge Landmark – (+) for outdoor positioning (-) does not cover indoor navigation scenarios
- Small Indoor Datasets
  - 7 Scenes – (+/-) highly textured scenes (-) only small areas (-) no global ambiguities
  - University – (-) ground truth obtained by using Google Tango; accuracy is between 6cm and 3m
- Large Indoor Datasets
  - Baidu – (+) high precision recording (-) only one route (-) low number of images (~600 training images)
  - Matterport3D – (+) differently textured scenes (-) no bigger environmental and illumination changes
  - Wijams – (+) differently textured scenes (-) no environmental and illumination changes
  - InLoc – (+) many environmental changes (-) only images in small office areas
  - TU Munich Large Scale-Indoor dataset – (-) no appearance changes such as moved structures or occlusion

Related Work

Evaluation Criteria for Camera-based Positioning Schemes

- Most of the existing evaluation methods have flaws
  - Overlapping Training/Test data, but **no systematic test approaches**
  - Methodically recorded training data, **randomly sampled test data**
  - Test **data selection usually is not described** sufficiently

- Accuracy metrics vary throughout different works
  - Most common is a percentage of predictions below a threshold (e.g. position error <5cm)
  - In some works the median position and orientation error is used
  - ➢ **Leaves out error over time, error distribution, axis-specific errors and location-specific behavior, i.e., problematic untextured walls**

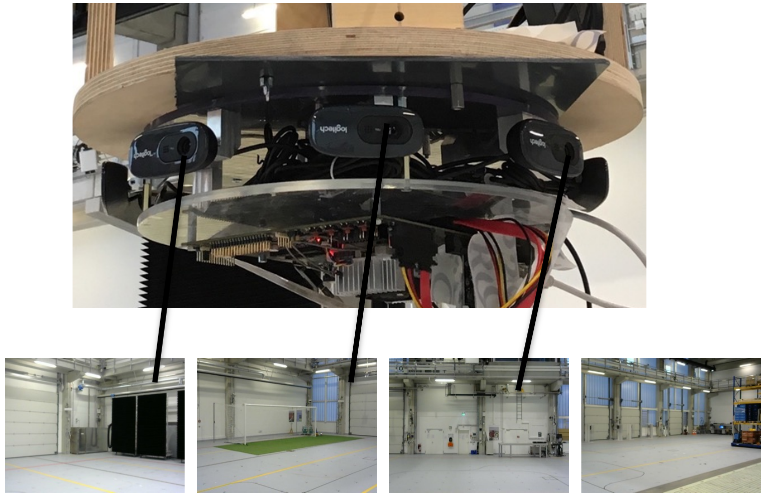
Fraunhofer

Our Contribution

- Compiling a large **dataset** in a realistic logistics environment with special test cases
  - For sufficient training and testing the setup evaluation criteria
- Standardized **evaluation criteria**
  - For tackling the problem of variating criteria in different works
- **Application** of the evaluation criteria and the dataset on a Deep Learning Architecture with more precise metrics

Warehouse - A dataset for applying standardized evaluation criteria

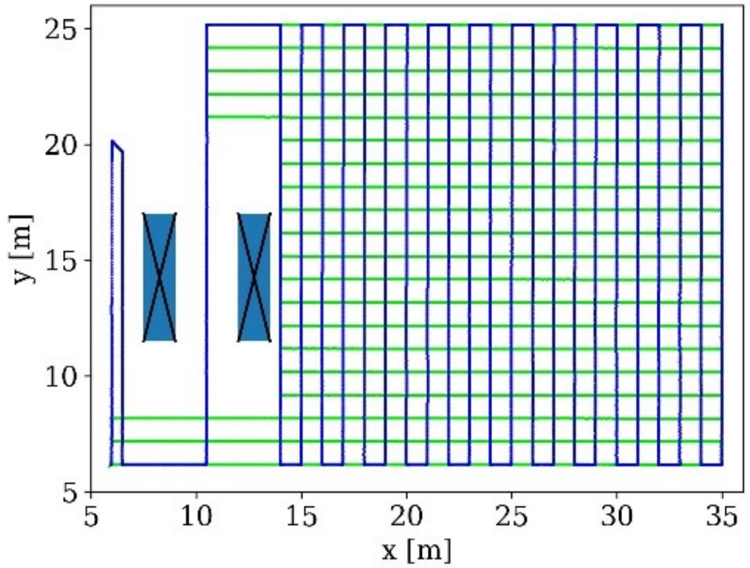
- ***Warehouse* dataset**
  - Aim: providing a solid basis for the development and evaluation of ML-based positioning schemes
  - Area: **1,320m^2**; Images: **464,804**; Image size: **640 x 480** pixels
  - Images labeled with a **sub-millimeter position** and **sub-degree orientation** using laser-based reference system (Nikon iGPS)
  - A Recording platform with 300mm diameter that carries **eight cameras** (calibrated Logitech C270) facing in different directions.
  - Recording took place in the Fraunhofer L.I.N.K. hall in Nuremberg
- The dataset provides **training** and **evaluation** sequences
  - 2 Training sequences to cover the area of warehouse
  - 8 Evaluation sequences for tackling the evaluation criteria

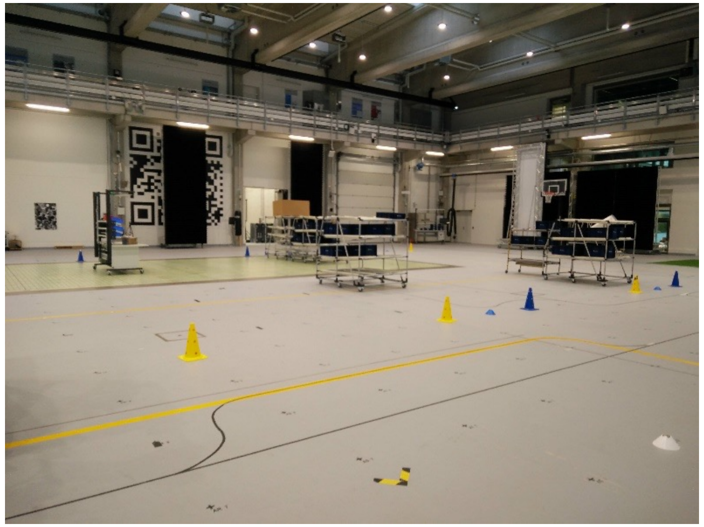


…

Warehouse - A dataset for applying standardized evaluation criteria

- **Training sequences**
  - Combination of horizontal and vertical grid
  - Grid spacing of 0.5m

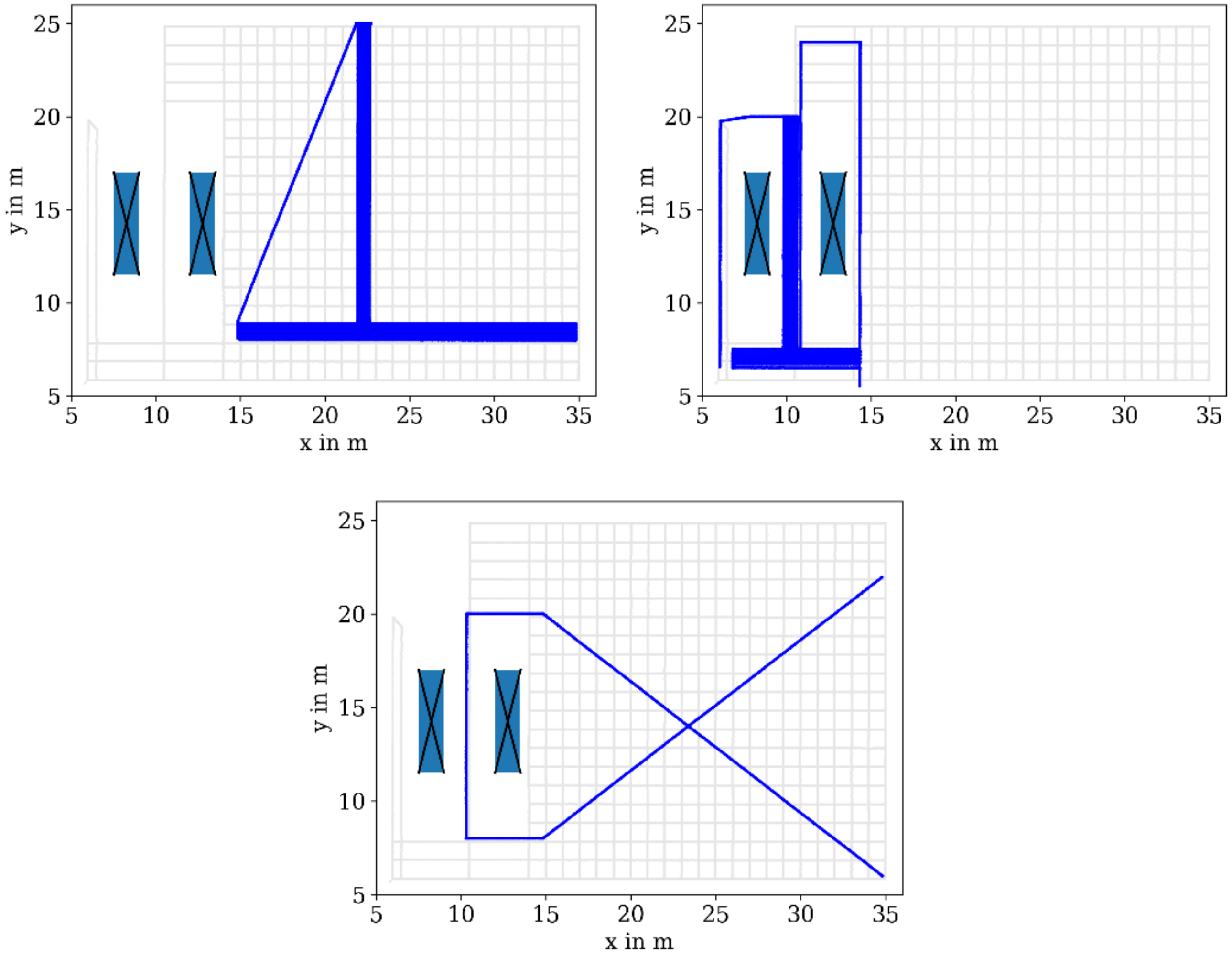




Fraunhofer

Evaluation Criteria and the according datasets

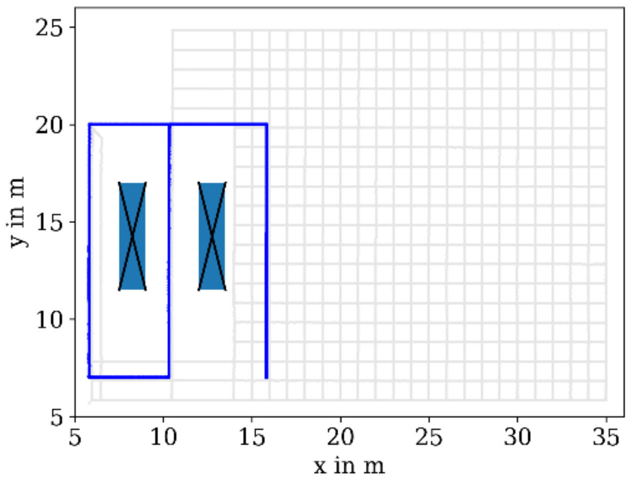
- **Generalization Criteria (1/6)**
  - Algorithm can predict previously unseen positions that are close to seen positions, but fails on areas further away
- 3 Datasets
  - Generalization large space (0.1m spacing)
    - Large open area with larger distances to more ambiguous global features
  - Generalization small space (0.1m spacing)
    - Only in close area around high racks
  - Cross
    - Varying distances than training

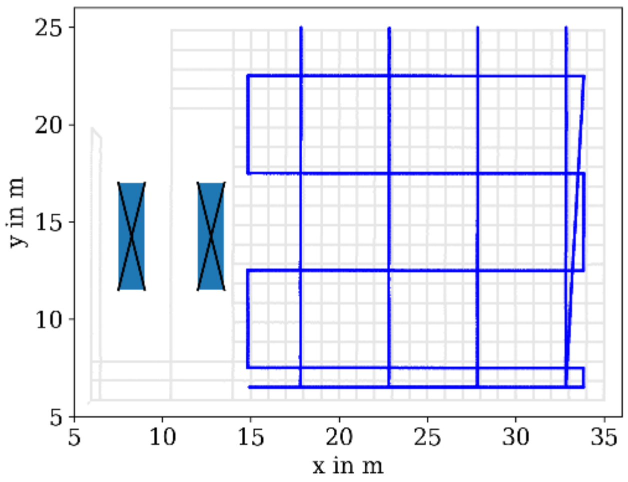


Fraunhofer

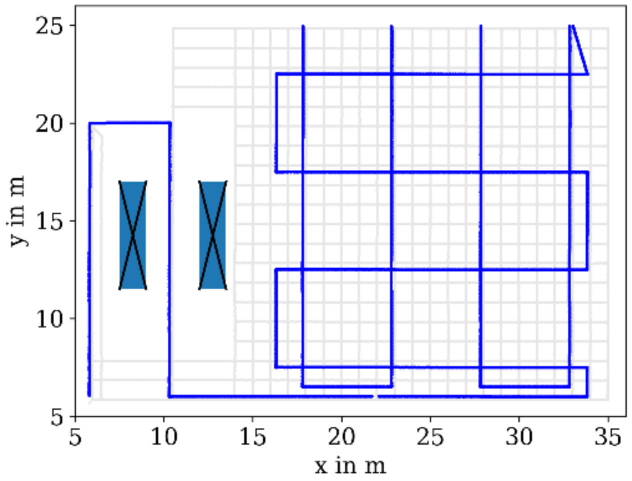
Evaluation Criteria and the according datasets

- **Environmental scaling (2/6)**
  - Positioning performance can differ over area scales
  - 2 datasets
    - Small scale area
    - Large scale area





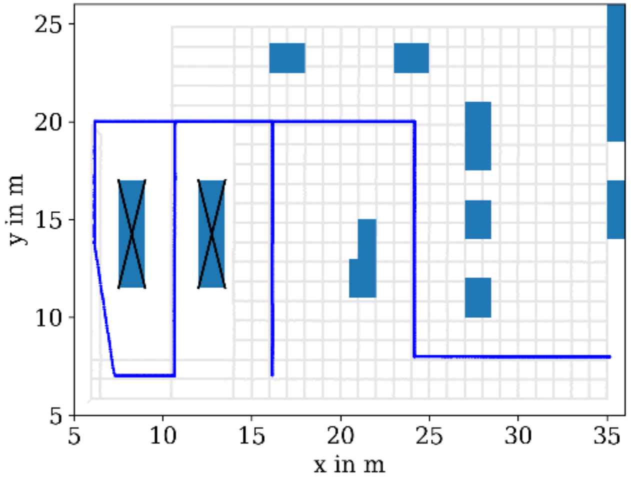
- **Scale Transition (3/6)**
  - Small and large scale areas in same dataset can affect performance (e.g. due to scale-invariant features)
  - 1 dataset
    - Scale transition

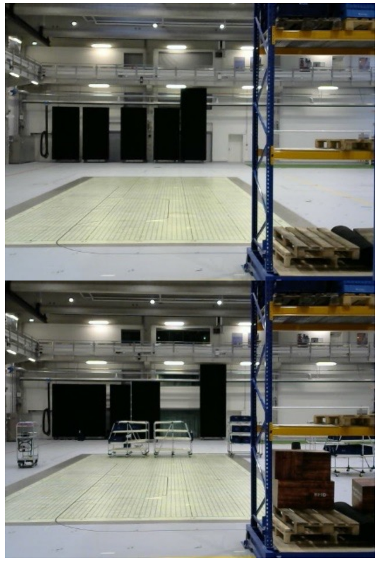


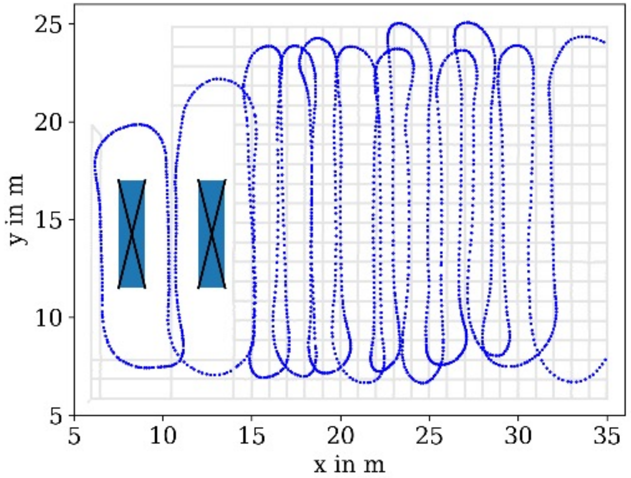
Fraunhofer

A dataset for applying standardized evaluation criteria

- **Volatility (4/6)**
  - Algorithm can fail to generalize to changed/volatile features which are not in the training dataset
  - 1 dataset
    - Volatility (new, movable racks are added to the scene)
- **Motion Artifacts (5/6)**
  - Blurry images, unsteady angles or new view points can influence the prediction performance heavily
  - 1 dataset
    - Forklift (recording platform on forklift)
- **Ambiguity (6/6)**
  - Ambiguous (i.e., repetitive or untextured) features in environment affect prediction accuracy



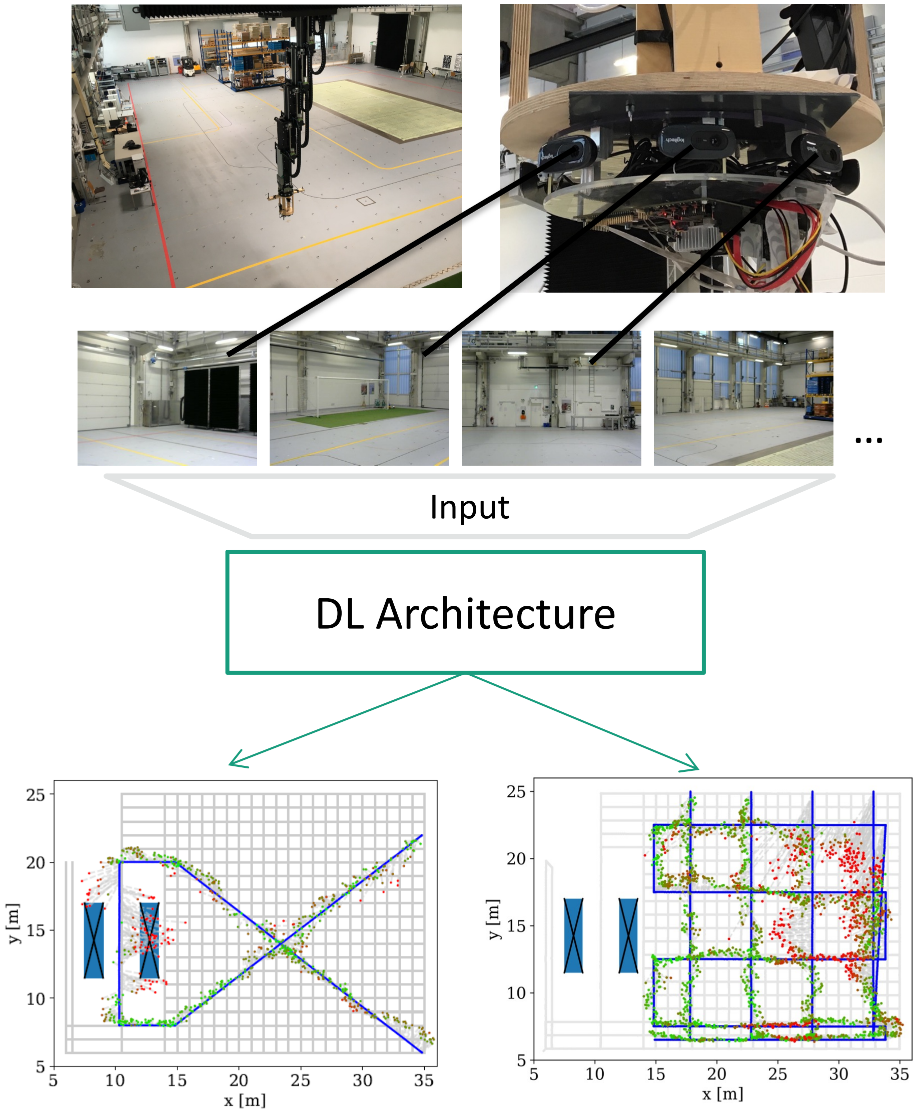




Fraunhofer

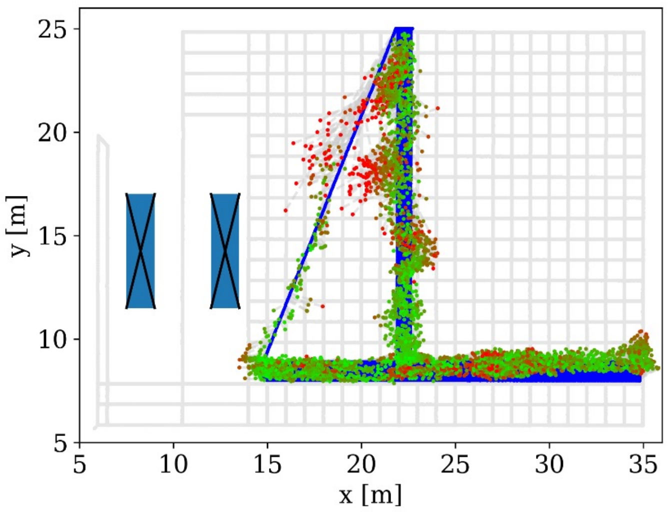
Applying the Evaluation Criteria to a DL positioning schemes

- Utilized Deep Learning system
  - **PoseNet** by Kendall et al. (2015) [2]
  - **Convolutional Neural Network** based on GoogLeNet for inferring positions from images
  - Training on Nvidia GTX 1080, Test on Nvidia GTX 1070
- **Metrics** for evaluating the DL system
  - Mean Absolute Error 2D (MAE 2D)
  - Circular Error Probable 2D (CEP)
  - Circular Error 95 Percentile 2D (CE95)
  - Rotation Circular Error Probable (RCEP)



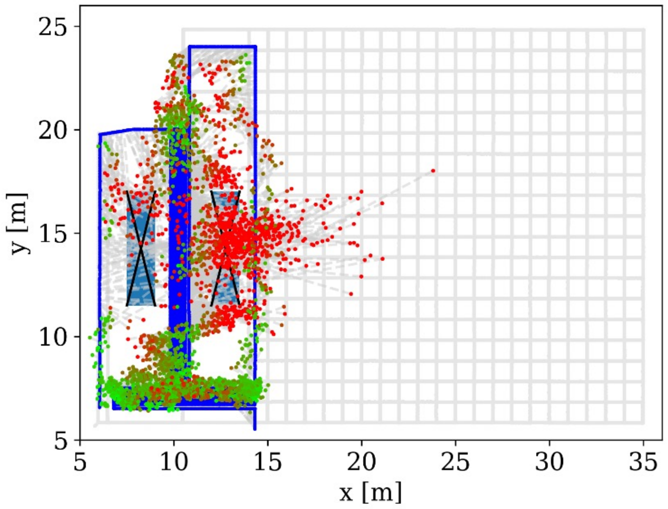
Results

2D Error < 2m   2D Error > 2m



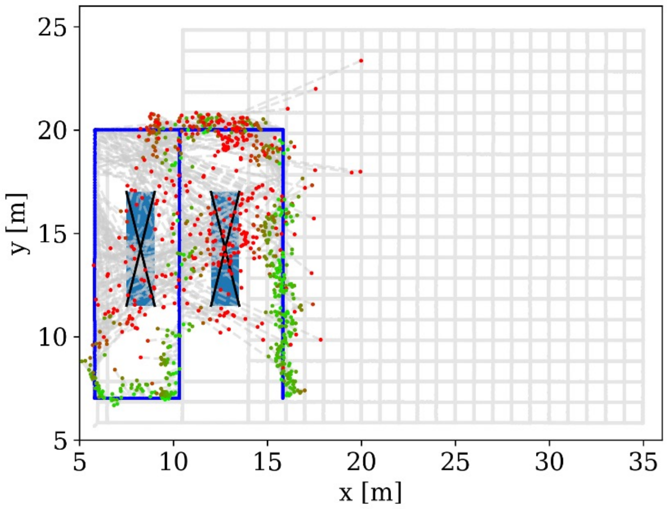
Generalization Open Area

MAE 2D	1.72m
CEP	1.06m
CE95	5.05m
RCEP	0.27°



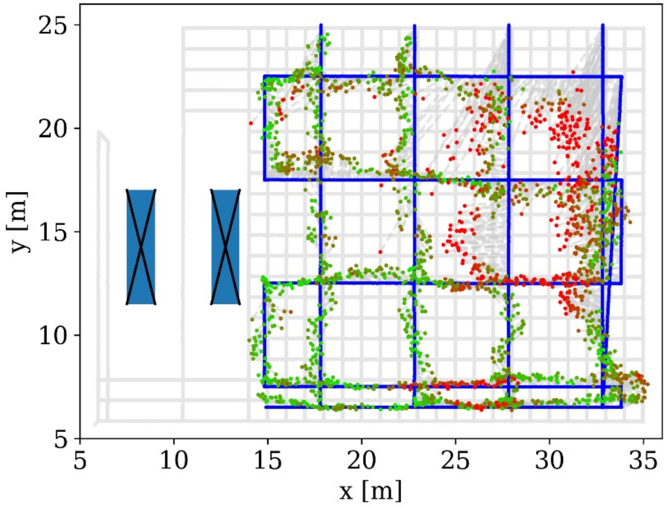
Generalization Rack Area

MAE 2D	2.43m
CEP	1.76m
CE95	7.96m
RCEP	0.457°



Small Scale

MAE 2D	2.31m
CEP	1.17m
CE95	8.99m
RCEP	0.18°

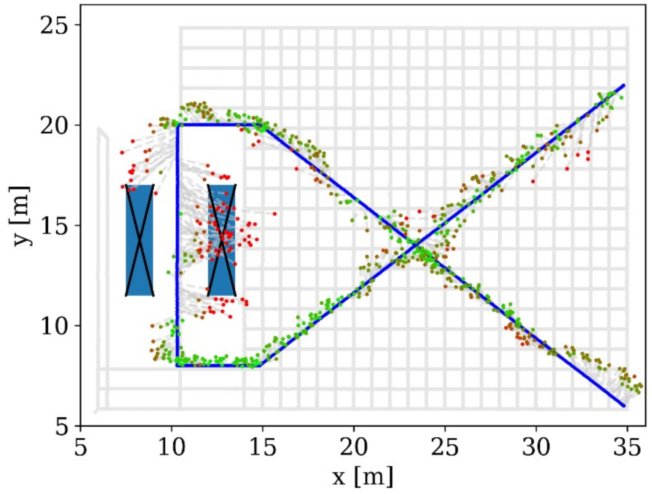


Large Scale

MAE 2D	1.14m
CEP	0.90m
CE95	2.83m
RCEP	0.18°

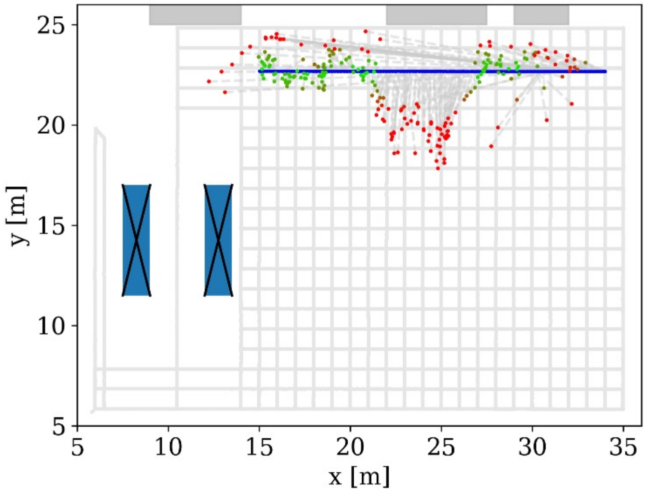
Results

2D Error < 2m   2D Error > 2m



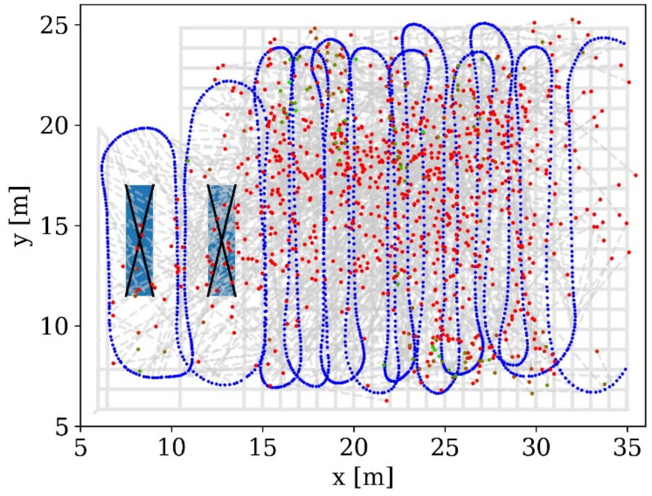
Cross

MAE 2D	1.08m
CEP	0.86m
CE95	3.08m
RCEP	0.18°



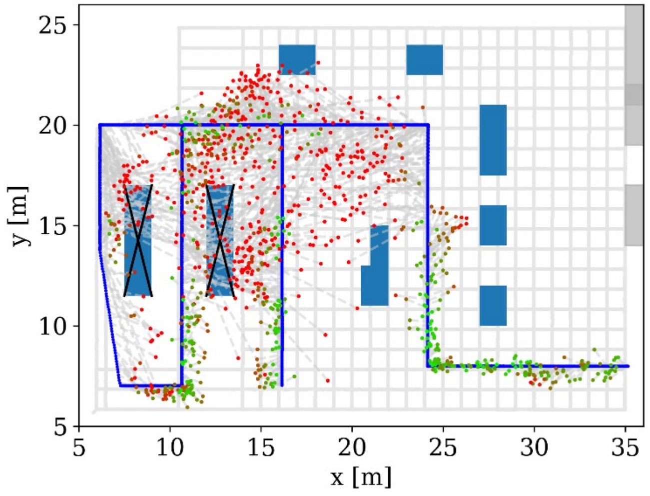
Ambiguity

MAE 2D	2.59m
CEP	1.26m
CE95	15.86m
RCEP	0.32°



Forklift Motion

MAE 2D	6.67m
CEP	5.42m
CE95	16.6m
RCEP	145.6°



Volatility

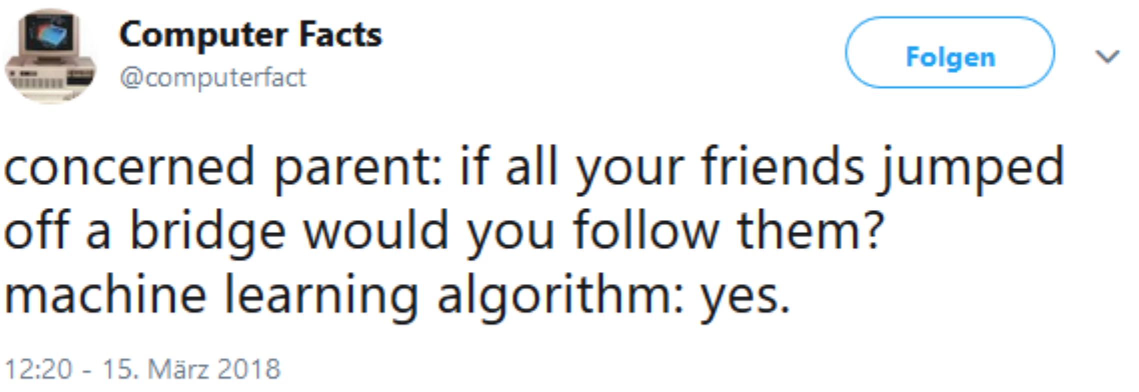
MAE 2D	2.5m
CEP	1.74m
CE95	7.35m
RCEP	0.56°

Fraunhofer

Conclusions

- **Contributions**
  - Introduction of the **first dataset for self-positioning** in a large **industrial indoor scenario** with high precision ground truth labels
  - Introduction of **six criteria** to properly **evaluate ML-based positioning schemes**
  - **Application** of the criteria and the dataset on a popular ML-based algorithm shows their **strengths** and **weaknesses** in different scenarios
  - ➢ Using our contributions, ML-based positioning systems can be evaluated and developed under consistent criteria with better insights and comparability
- **Future Work**
  - Use our dataset to evaluate complementary positioning schemes
    - …that incorporate **temporal correlation of features** using, e.g. Bayesian filters and LSTM cells
    - …that incorporate **additional sensors**, e.g. IMUs, etc.

Questions?



- *Warehouse* dataset download
  - https://www.iis.fraunhofer.de/warehouse

- Contact
  - Christoffer Löffler – christoffer.loeffler@iis.franhofer.de
  - Sascha Riechel – riechesa@iis.fraunhofer.de

Fraunhofer

References

- R. Mautz and S. Tilch, “Survey of optical indoor positioning systems,” in Proc. IEEE Intl. Conf. Indoor Positioning and Indoor Navigation, (Guimares, Portugal), pp. 1–7, 2011.
- A. Kendall, M. Grimes, and R. Cipolla, “Posenet: A convolutional network for real-time 6-dof camera relocalization,” in Proc. IEEE Intl. Conf.ComputerVision,(SantiagodeChile,Chile),pp.2938–2946,2015.
- Z. Laskar, I. Melekhov, S. Kalia, and J. Kannala, “Camera relocalization by computing pairwise relative poses using convolutional neural network,” in Proc. IEEE Intl. Conf. Computer Vision Workshop, (Santiago de Chile, Chile), pp. 920–929, 2017.
- J. Shotton, B. Glocker, C. Zach, S. Izadi, A. Criminisi, and A. Fitzgibbon, “Scene coordinate regression forests for camera relocalization in rgb-d images,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, (Portland, OR), pp. 2930–2937, 2013.
- X. Sun, P. Xie, Luo, and L. Wang, “A dataset for benchmarking imagebased localization,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, (Honolulu, HI), pp. 5641–5649, 2017.
- A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Niessner, S. Savva, A. Zeng, and Y. Zhang, “Matterport3d: Learning from rgb-d data in indoor environments,” in Proc. Intl. Conf. 3D Vision, (QingDao, China), 2017.
- E. Wijmans and Y. Furukawa, “Exploiting 2d floorplan for buildingscale panorama rgbd alignment,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, (Honolulu, HI), pp. 308–316, 2017.
- H. Taira, M. Okutomi, T. Sattler, M. Cimpoi, M. Pollefeys, J. Sivic, T. Pajdla, and A. Torii, “Inloc: Indoor visual localization with dense matching and view synthesis,” in IEEE Conf. Computer Vision and Pattern Recognition, (Salt Lake City, UT), 2018.
- F. Walch, C. Hazirbas, L. Leal-Taix´e, T. Sattler, S. Hilsenbeck, and D. Cremers, “Image-based localization using lstms for structured feature correlation,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, (Honolulu, HI), pp. 627–637, 2017.
- Hammad Tariq, “Blog Writing - Introduction to ERP Systems - Using SAP”, URL: https://www.researchgate.net/project/Blog-Writing-Introduction-to-ERP-Systems-Using-SAP