# ACE-OPS Project

Changhao Chen, Lecturer

College of Intelligence Science

National University of Defense Technology

20th Jan 2021

# Background: Emergency Response

London Grenfell Tower fire
14 June 2017

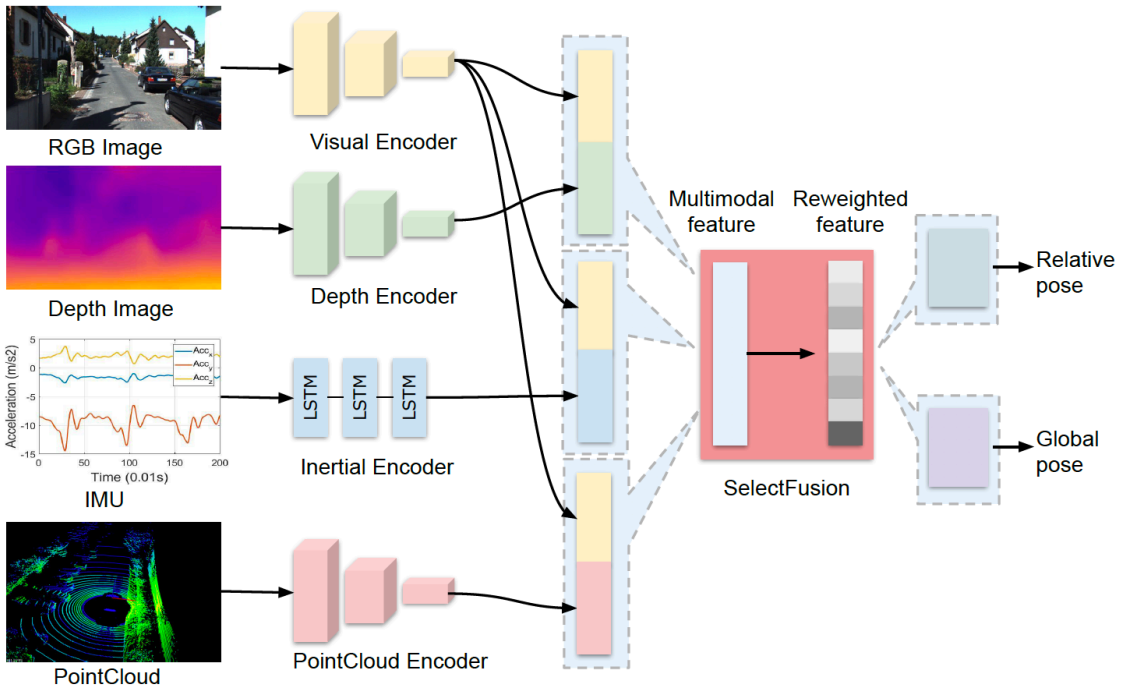
# Research Question

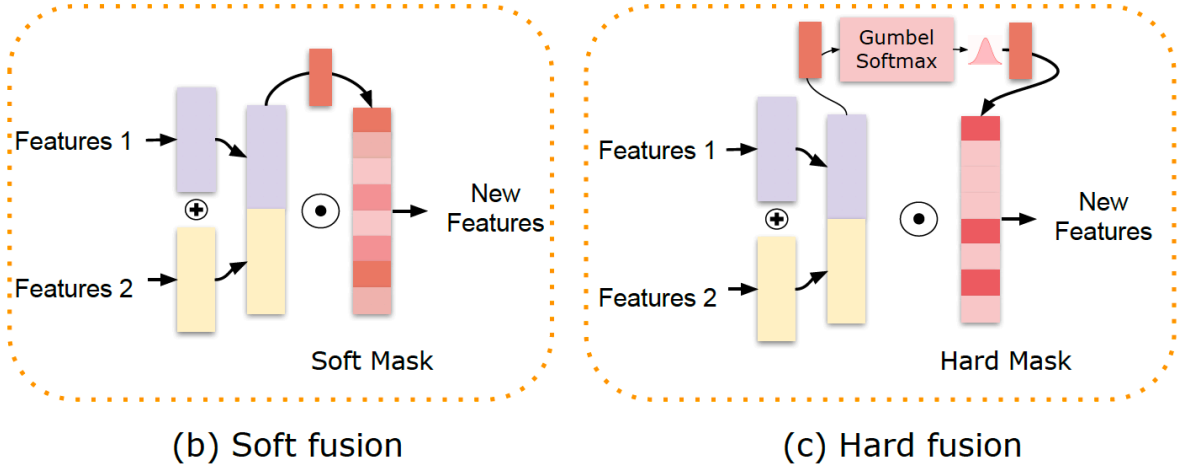
Can we develop learning methods using multimodal data to achieve accurate and robust localization without hand-crafted engineering?

(a) Effective Sensor Fusion
(b) Localization with respect to existing map
(c) Self-supervised Learning of egomotion

C. Chen et al. (2020) A Survey on Deep Learning for Localization and Mapping: Towards the Age of Spatial Machine Intelligence. Submitted to TPAMI

# A Generic framework of Selective Sensor Fusion

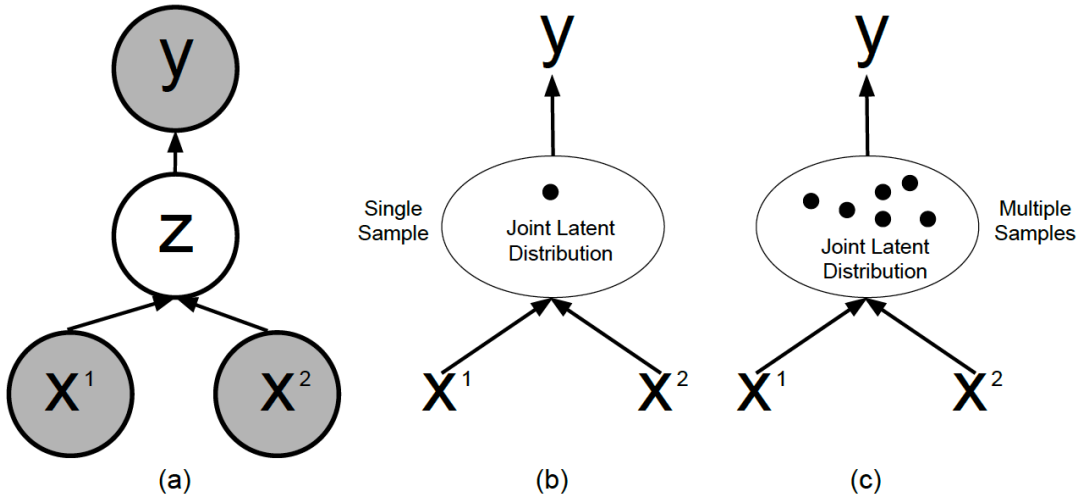



- A general framework for sensor fusion that selects useful features for localization
- Applied in depth-vision relocalization, lidar-vision odometry, and visual-inertial odometry

C. Chen et al. (2020) Learning Selective Sensor Fusion for States Estimation. Submitted to TNNLS

# VMLoc: Multimodal Variational Sensor Fusion



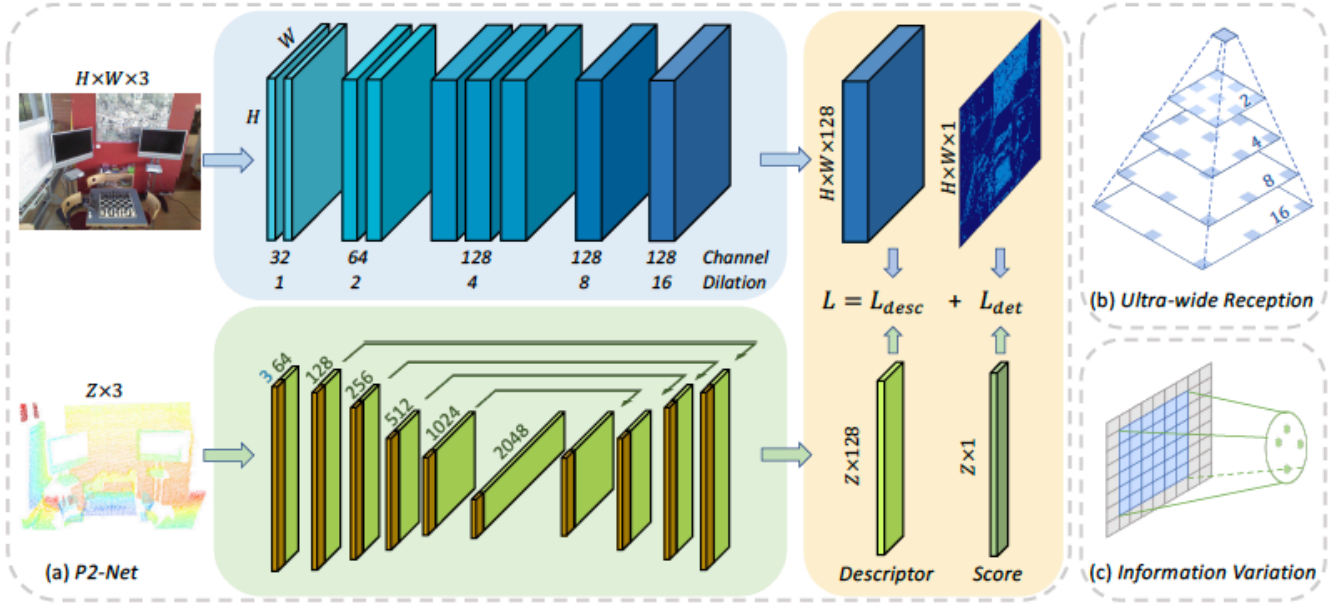

| Scene | MapNet(V) | AtLoc(V) | VMLoc(V,D) |
|---|---|---|---|
| LOOP1 | $8.76m, 3.46°$ | $8.61m, 4.58°$ | **7.70m, 3.23°** |
| LOOP2 | $9.84m, 3.96°$ | $8.86m, 4.67°$ | **7.76m, 3.16°** |
| FULL1 | $41.4m, 12.5°$ | $29.6m, 12.4°$ | **19.5m, 4.32°** |
| FULL2 | $59.3m, 14.8°$ | $48.2m, 11.1°$ | **35.2m, 8.99°** |
| Average | $29.8m, 8.68°$ | $23.8m, 8.19°$ | **17.5m, 4.92°** |

| Scene | image VMLoc | attention VMLoc | PoE VMLoc | VMLoc |
|---|---|---|---|---|
| LOOP1 | $8.60m, 4.57°$ | $9.16m, 4.96°$ | $8.57m, 3.98°$ | **7.70m, 3.23°** |
| LOOP2 | $8.50m, 3.90°$ | $9.78m, 5.66°$ | $8.99m, 3.79°$ | **7.76m, 3.16°** |
| FULL1 | $30.1m, 10.8°$ | $31.2m, 6.04°$ | $30.0m, 7.54°$ | **19.5m, 4.32°** |
| FULL2 | $48.1m, 9.61°$ | $46.5m, 10.1°$ | $45.9m, 10.5°$ | **35.2m, 8.99°** |
| Average | $23.9m, 7.22°$ | $19.3m, 5.35°$ | $18.7m, 5.16°$ | **17.5m, 4.92°** |

- Variational inference based multimodal fusion strategy
- Our multimodal relocalization outperforms other baselines

R. Zhou et al. (2020) VMLoc: Variational Fusion For Learning-Based Multimodal Camera Relocalization. Accepted by AAAI 2021

# P2-Net: Learning 2D-to-3D descriptors matching



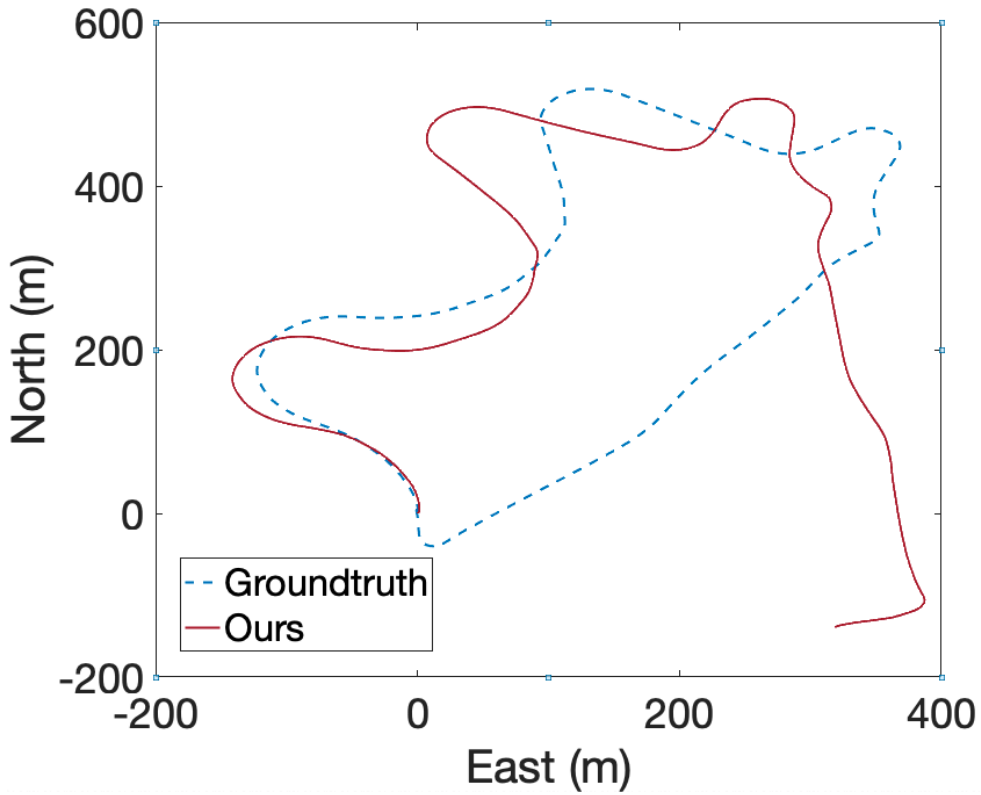

| | Feature Matching Recall | | | | | | |
|---|---|---|---|---|---|---|---|
| SIFT + SIFT3D | Not Match | | | | | | |
| D2_Triplet | Not Converge | | | | | | |
| D3_Contrastive | Not Converge | | | | | | |
| R2D2_Based | 95.1 | 97.3 | **100** | 89.4 | 91.1 | 88.7 | 16.2 |
| ASL_Based | 95.3 | 96.0 | **100** | 34.3 | 41.6 | 47.5 | 11.9 |
| P2[w/o Det] | 78.0 | 82.0 | 99.1 | 48.5 | 64.7 | 43.8 | 15.0 |
| P2[Mixed] | 82.5 | 95.5 | 99.7 | 86.4 | 70.5 | 83.8 | 16.1 |
| P2[D2_Det] | **100** | 99.7 | **100** | 93.6 | 98.4 | 94.0 | 74.3 |
| P2[D3_Det] | 99.0 | 99.7 | **100** | 83.8 | 68.0 | 78.4 | 17.8 |
| P2[Rand] | **100** | 99.6 | 99.8 | 90.8 | 83.2 | 82.5 | 14.3 |
| P2-Net | **100** | **100** | **100** | **97.3** | **98.5** | **96.3** | **88.8** |

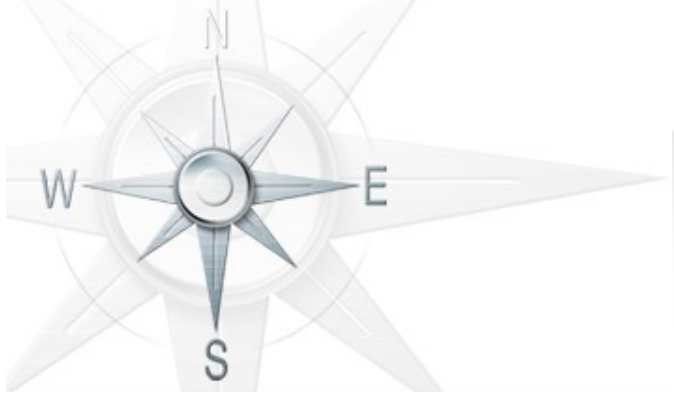
- Learning based 2D-to-3D features matching.
- Applied in map matching (2D scan to 3D points map), global relocalization.
- Propose P2-Net, a novel framework, including 2D features extractors, 3D features extractors, and novel descriptors learning and matching strategy

B. Wang et al. (2020) P2-Net: Joint Description and Detection of Local Features for Pixel and Point Matching. Submitted to CVPR 2021

# Ongoing work: Self-supervised Localization In The Wild

- Self-supervised learning of egomotion from cameras and IMUs
- Camera intrinsics are learned (No calibration is needed).
- Egomotions are scale consistent
- Wer are working to further improve model performance






UNIVERSITY OF OXFORD

# Thanks for your attention!

Changhao Chen

National University of Defense Technology

Email: chenchanghao@nudt.edu.cn