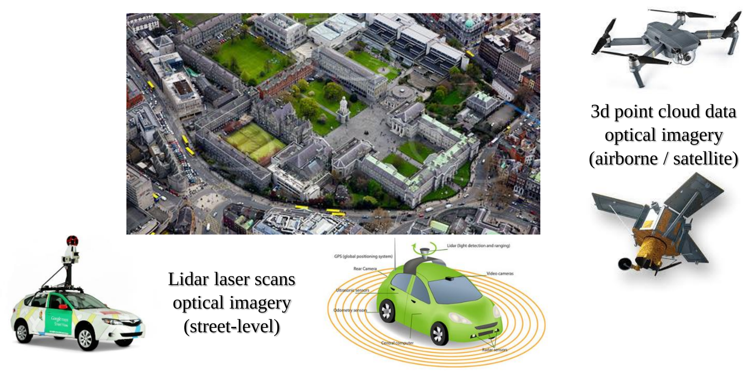
# OBJECT GEOLOCATION USING MRF BASED MULTI-SENSOR FUSION

*Vladimir A. Krylov* and Rozenn Dahyot
ADAPT Centre, School of Computer Science & Statistics, Trinity College Dublin, Ireland

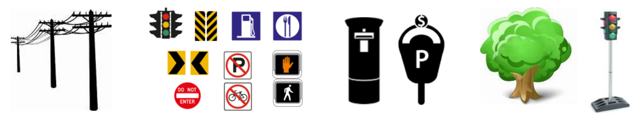
Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin
Engaging Content
Engaging People

## Objectives

- Automatic complex scenes analysis: multiple objects.
- Recycling abundant existing image datasets.
- Efficient detail-preserving fusion of multi-sensor data.




- We design a multi-sensor fusion pipeline that can perform discovery and geolocation of objects across various imaging and sensory modalities.
- In this work we study the case of two modalities: **Street Level Imagery** and **LiDAR.**
- We consider detection of compact stationary objects:



## Image Processing

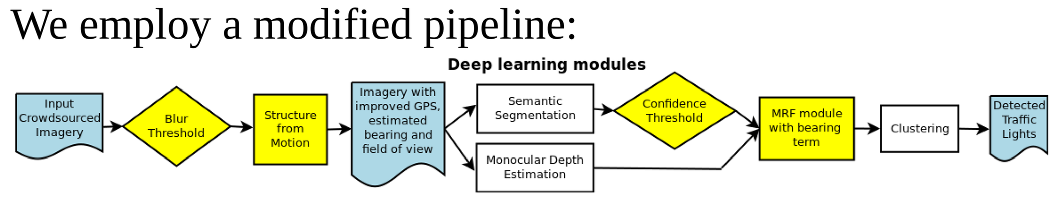
- **Assumptions**:
  - Street level imagery is the primary source.
  - Object sparsity: 1m apart.
- **Street Level Imagery**: We perform
  - ✓ Semantic Segmentation
  - ✓ Monocular depth estimation

  via state-of-the-art fully CNN models.
  These are finetuned on *Cityscapes* and *Mapillary Vistas* street level imagery datasets.
- **LiDAR data**: We obtain
  - ✓ Candidate point extraction

  via template matching.

## Crowdsourced Imagery

In [3] we use crowdsourced *Mapillary* images for object discovery and geolocation.
We employ a modified pipeline:




We rely on *Laplacian filtering* for thresholding and *Structure from Motion* to estimate bearing, adjust GPS.

**Conclusions**:

- Requires more data to achieve comparable recall;
- Reduced position estimation accuracy.

## Multi-Sensor Fusion Pipeline

- We perform optimization for object discovery and geolocation based on the following input estimates:
  - ✓ Individual discovered objects (semantic segmentation)
  - ✓ Monocular depth estimated for each discovered object $\Delta_i$
  - ✓ LiDAR candidate matches

- We define a **Markov Random Field** (MRF) model over the space $X$ of all view-rays intersections:
  - label $z=0$ if not occupied by object
  - label $z=1$ if occupied
- MRF configuration is characterized by its corresponding energy $U$. *Optimal = minimum of U.* Optimization: *ICM*.

**Energy terms**:

$\Delta$ – depth estimates
$d$ – triangulated distances
$L$ – distance to LiDAR match
$R_i$ - subset of $X$ on the same ray with $x_i$

- Unary consistency terms:

$$u_D(z_i|\mathcal{X},\mathcal{Z}) = z_i \sum_{j=1,2} \|\Delta_{ij} - d_{ij}\|^2$$

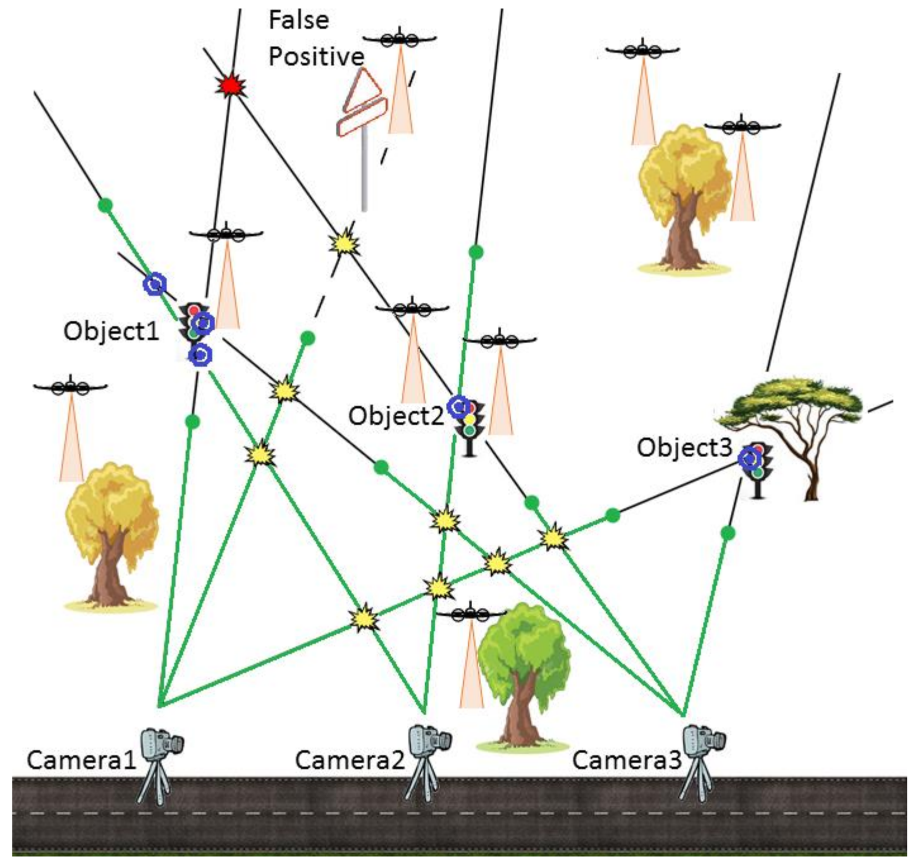
$$u_L(z_i|\mathcal{X},\mathcal{Z}) = z_i L_i^2$$

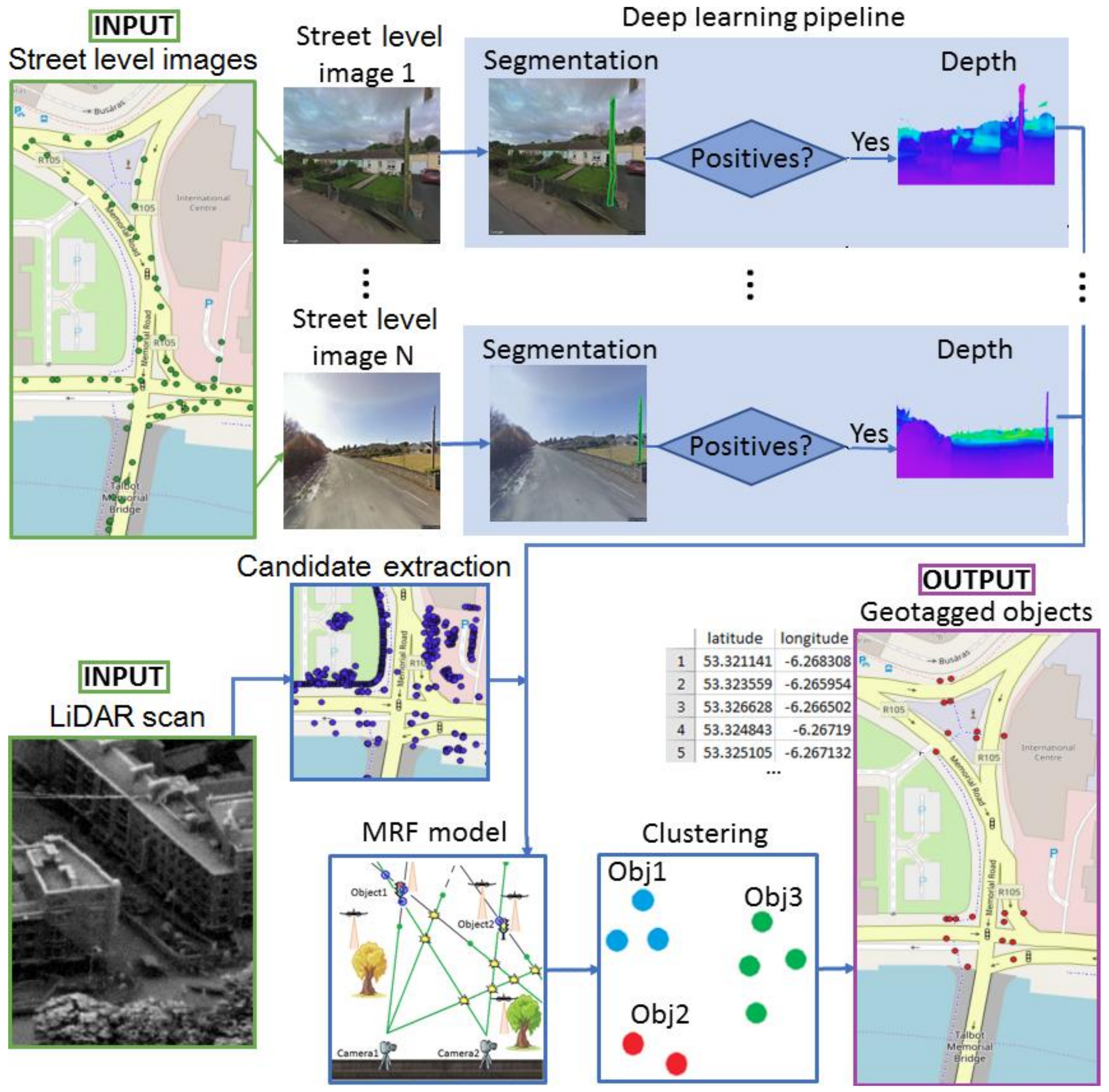
- Pairwise term. *No occlusions. No spread.*

$$u_0(R_i|\mathcal{X},\mathcal{Z}) = \prod_{x_n \in R_i} (1 - z_n)$$

- High-order term. *Penalize not matched rays.*

$$u_C(R_i|\mathcal{X},\mathcal{Z}) = \sum_{x_m, x_n \in R_i} z_m z_n \|x_m - x_n\|^2$$



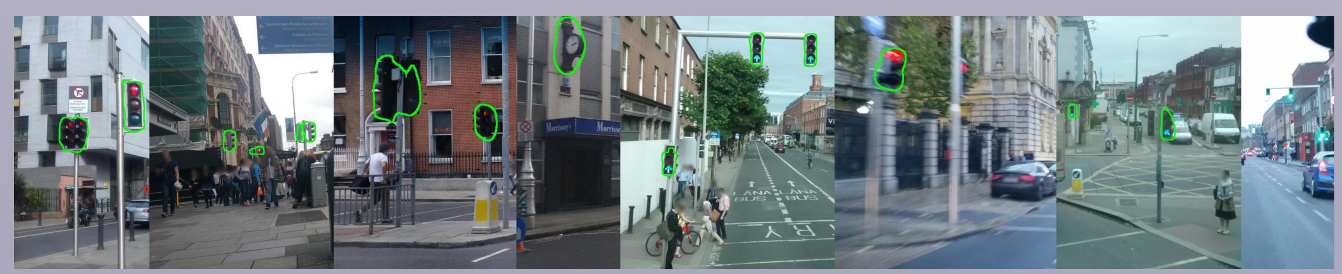




**Total energy**:

$$\mathcal{U}(\mathbf{z}) = \sum_{\forall x_i \in \mathcal{X}} \left[ c_D\, u_D(z_i) + c_L\, u_L(z_i) \right] + \sum_{\forall \text{ rays } R_j} \left[ c_C\, u_C(R_j) + c_0\, u_0(R_j) \right],$$

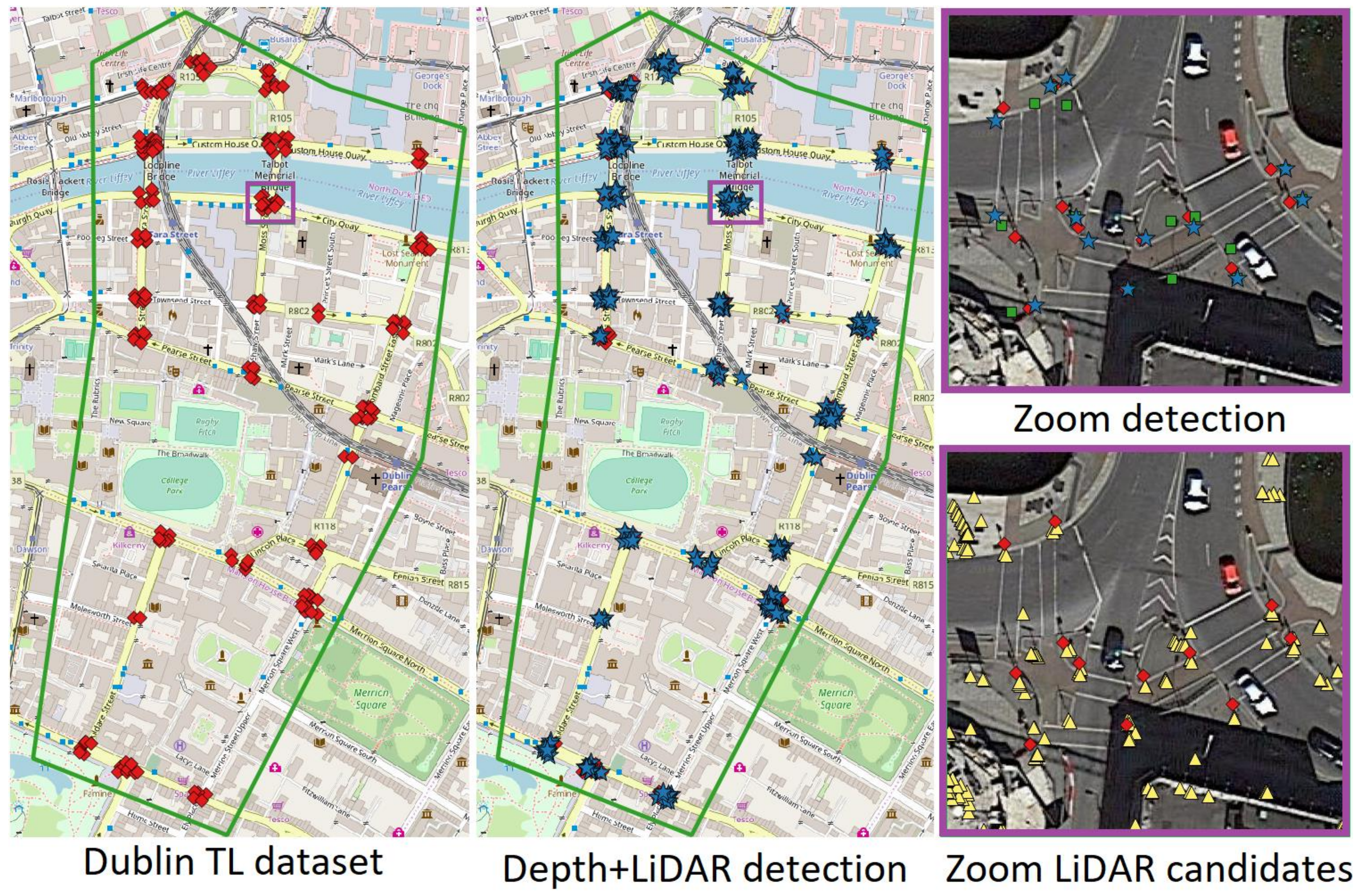
$$c_D + c_L + c_C + c_0 = 1$$

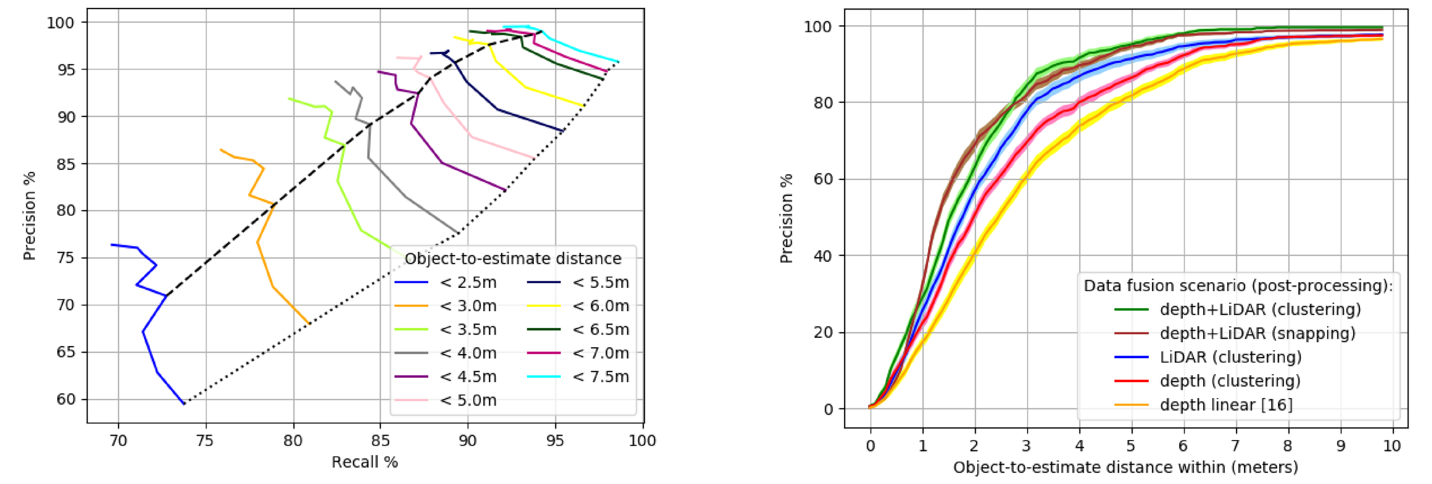
- Post-processing strategies:
  - *Clustering* via averaging (lower precision);
  - *Snapping* to LiDAR positions (lower recall).
- The same MRF-based architecture can incorporate:
  - Further image and data modalities;
  - Geometrical and position assumptions;
  - SLAM-like formulations.



## Experimental Results

- We introduce a **dataset** with 209 pole-mounted traffic lights in central Dublin.
- We employ:
  - ✓ **Google Street View** imagery: 1307 panoramas
  - ✓ **Airborne LiDAR** scan reporting 12300 pole-like matches




**an estimate is considered true positive if it is within m meters of a ground truth point.*

## Conclusion

**Fully automated** | **Customisable to various objects** | **Efficient multi-sensor fusion**

Relies on Street Level Imagery as primary detection source: performance validated on expert (*Google Street View*) and crowdsourced (*Mapillary*) imagery.

## References

[1] V. A. Krylov, E. Kenny, and R. Dahyot, "Automatic discovery and geotagging of objects from street view imagery," Remote Sens., vol. 10, no. 5, 2018.

[2] V. A. Krylov, R. Dahyot. "Object Geolocation using MRF-based Multi-sensor Fusion", Int'l Conf. Image Proc. (ICIP) 2018, Proc. ICIP 2018, pp. 2745-2749, 2018.

[3] V. A. Krylov, R. Dahyot. "Object Geolocation from Crowdsourced Street Level Imagery", European Conference on Machine Learning (ECML) Workshops, 2018.

The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

Fondúreacht Eolaíochta Éireann
Science Foundation Ireland
Ireland's European Structural and Investment Funds Programmes 2014-2020
Co-funded by the Irish Government and the European Union
European Union
European Regional Development Fund