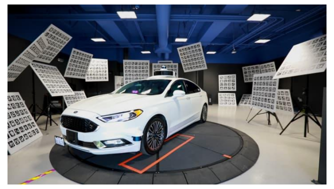
UniCal: Unified Neural Sensor Calibration

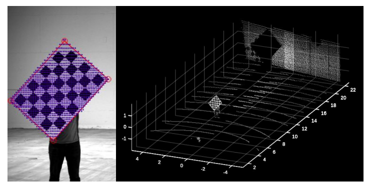
Ze Yang*, George Chen*, Haowei Zhang, Kevin Ta, Ioan Andrei Bârsan,
Daniel Murphy, Sivabalan Manivasagam, Raquel Urtasun

https://waabi.ai/unical

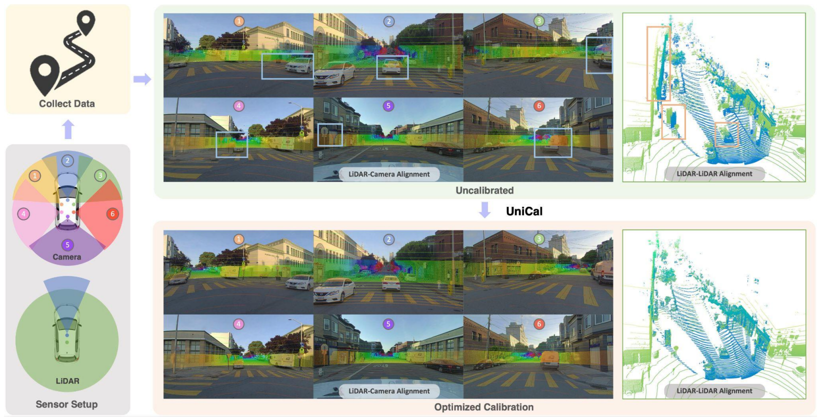
Motivation: Drive-and-Calibrate

+ Self-driving vehicles (SDVs) require accurate multi-sensor calibration
+ Existing approaches:
  - **Require Large Infrastructure**
  - **High Operational Effort**



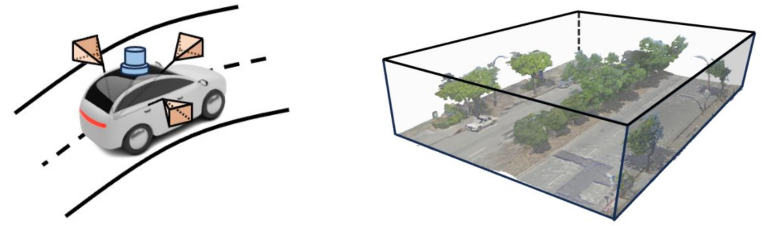


+ Goal: "drive and calibrate" - no targets, automatic, all jointly optimized
+ Approach:
  - Optimize sensor calibration with unified neural rendering framework
  - Enhance neural rendering specifically for multi-sensor calibration



Task Setting and Scene Representation

+ **Sensor Calibration Graph**
  - Optimizes sensor extrinsics w.r.t. a vehicle reference (IMU)
  - Assumes the intrinsics (e.g., focal length) are provided
  - Assumes vehicle trajectories (localization) are provided



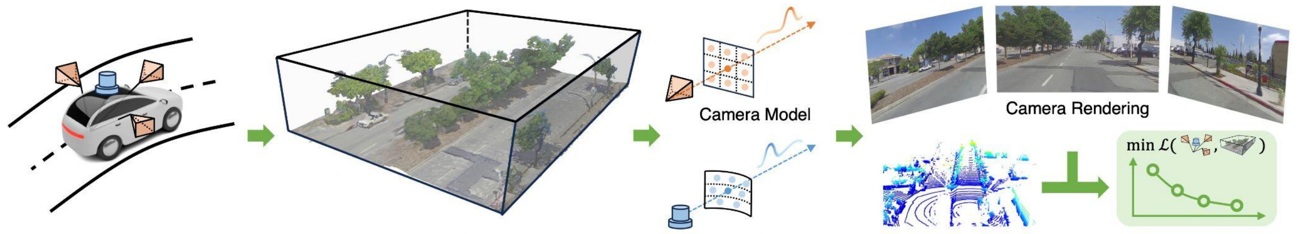
+ **Implicit Neural Scene Representation**
  - Multi-resolution feature grid with MLP networks

$$s, \mathbf{f} = \mathcal{Q}(\mathbf{x}) = \mathtt{MLP}(\{\mathtt{interp}(\mathbf{x}, \mathbf{G}^l)_{l=1}^{L}\})$$

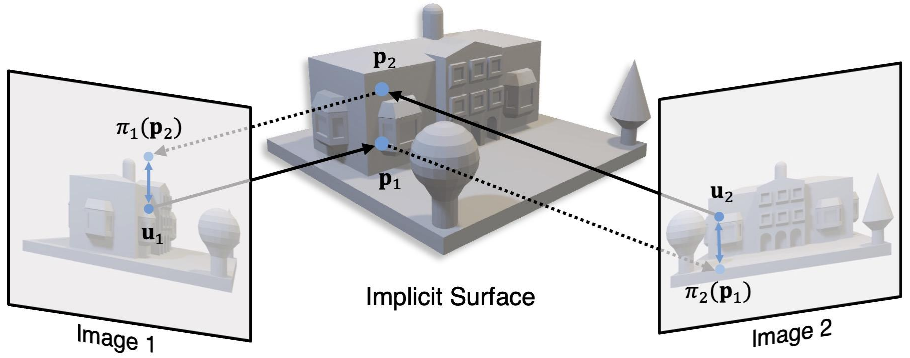
  - Generates camera and LiDAR though differentiable volume rendering

UniCal Method

+ **Overview**
  - Build the sensor calibration graph and implicit scene representation
  - Render camera and LiDAR with differentiable sensor models
  - Jointly optimize the multi-sensor extrinsics and underlying scene representation
  - Minimize the photometric and geometric consistency losses on collected outdoor data



+ **Surface Alignment Constraints**
  - Inferred correspondences should align with the underlying scene surfaces
    - Match image features
    - Raycast to 3D points
    - Project back to image
    - Compute pixel distance between points
  - Minimizing surface alignment distance ensures consistency across sensors and views



+ **Learning Sensor Calibration**
  - Photometric Consistency Loss (Image RGB and LiDAR intensity)

$$\mathcal{L}_{\text{rgb}} = \sum_{i=1}^{N_{\text{cam}}} \sum_{j=1}^{N} \left\| \hat{\mathbf{I}}_{\text{cam}}(\mathbf{u}_j^i | \mathbf{E}_{\text{cam}}^i) - \mathbf{I}_{\text{cam}}(\mathbf{u}_j^i) \right\|_2 \qquad \mathcal{L}_{\text{int}} = \sum_{i=1}^{N_{\text{lidar}}} \sum_{j=1}^{N} \left\| \hat{\mathbf{I}}_{\text{lidar}}(\mathbf{w}_j^i | \mathbf{E}_{\text{lidar}}^i) - \mathbf{I}_{\text{lidar}}(\mathbf{w}_j^i) \right\|_2$$

  - Geometric Consistency Loss (Image surface alignment and LiDAR depth)

$$\mathcal{L}_{\text{depth}} = \sum_{i=1}^{N_{\text{lidar}}} \sum_{j=1}^{N} \left\| \hat{D}(\mathbf{w}_j^i | \mathbf{E}_{\text{lidar}}^i) - D(\mathbf{w}_j^i) \right\|_1 \qquad \mathcal{L}_{\text{align}} = \sum_{i=1}^{N_{\text{cam}}} \sum_{j=1}^{N_{\text{cam}}} \sum_{k=1}^{M} \frac{\ell_{\text{surf}}\left(\mathbf{u}_k^i, \mathbf{u}_k^j | \mathbf{E}_{\text{cam}}^i, \mathbf{E}_{\text{cam}}^j\right)}{M}$$

  - Regularization Term (surface weight concentration, SDF satisfy Eikonal equation)

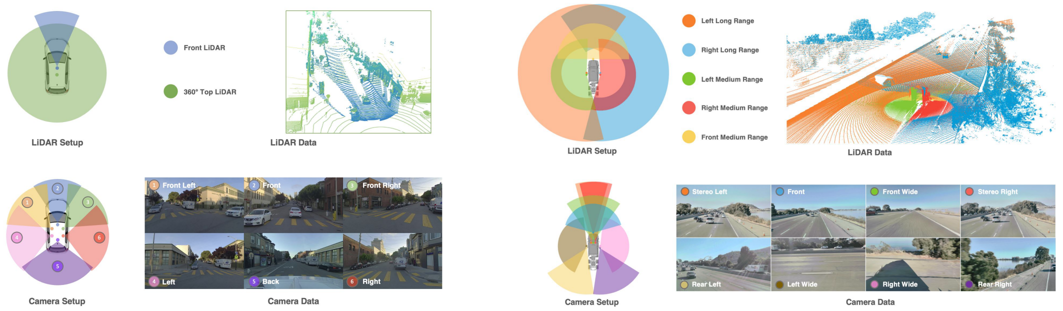
$$\mathcal{L}_{\text{reg}} = \sum_{\tau_i > \epsilon} \|w_i\|_2 + \lambda \sum_{\tau_i < \epsilon} (\|\nabla s(\mathbf{x}_i)\|_2 - 1)^2$$

  - Ray Sampling Priors (coarse-to-fine sampling strategy)

$$p(\beta) \propto h_{\text{min}} + \beta \cdot \mathtt{GaussBlur}(h, k_\beta)$$

Results

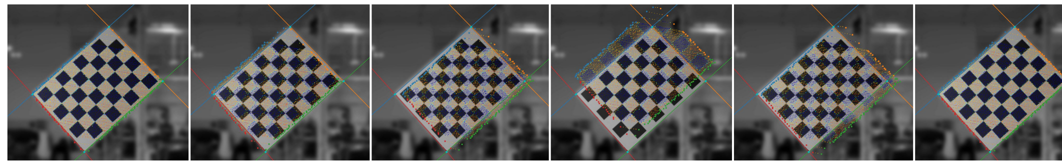
+ **Datasets**



PandaSet: Minivan

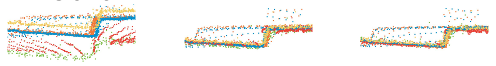
MS-Cal: Class-8 Truck

+ **Qualitative comparison with SOTA approaches**
  - Checkerboard Re-projection Comparison



  - Outdoor LiDAR-Camera Re-projection Comparison

  - Outdoor LiDAR-LiDAR Registration Comparison
    - Flat Ground
    - Curb



  - Rendering Comparison

+ **Quantitative comparison with SOTA approaches**

Methods	Sensors	Multi-Sensor Alignment		Camera Pose Accuracy		LiDAR Pose Accuracy		Rendering Quality			
		LiD↔Cam↓	LiD↔LiD↓	Rotation↓	Transl↓	Rotation↓	Transl↓	PSNR↑	SSIM↑	LPIPS↓	Depth↓
Point-to-Plane ICP [7]	LiD-LiD	-	**2.811**	-	-	0.043	0.014	-	-	-	-
MLCC [36]	LiD-LiD	-	3.064	-	-	0.068	0.022	-	-	-	-
Mutual Info [43]	LiD-Cam	39.76	-	1.182	0.200	-	-	-	-	-	-
Edge and Plane [70]	LiD-Cam	9.83	-	0.358	0.030	-	-	-	-	-	-
Pose Graph Optim [71]	Full	11.14	2.962	0.438	**0.029**	0.049	0.012	30.11	0.854	0.422	0.073
SC-NeRF [23]	Cam-Cam	58.80	-	0.544	0.204	-	-	29.82	0.854	0.438	-
INF [72]	LiD-Cam	47.27	8.996	0.645	0.189	0.368	0.116	30.77	0.872	0.400	0.148
Ours	Full	**9.27**	2.847	**0.186**	0.033	**0.036**	**0.008**	**31.96**	**0.903**	**0.344**	**0.035**