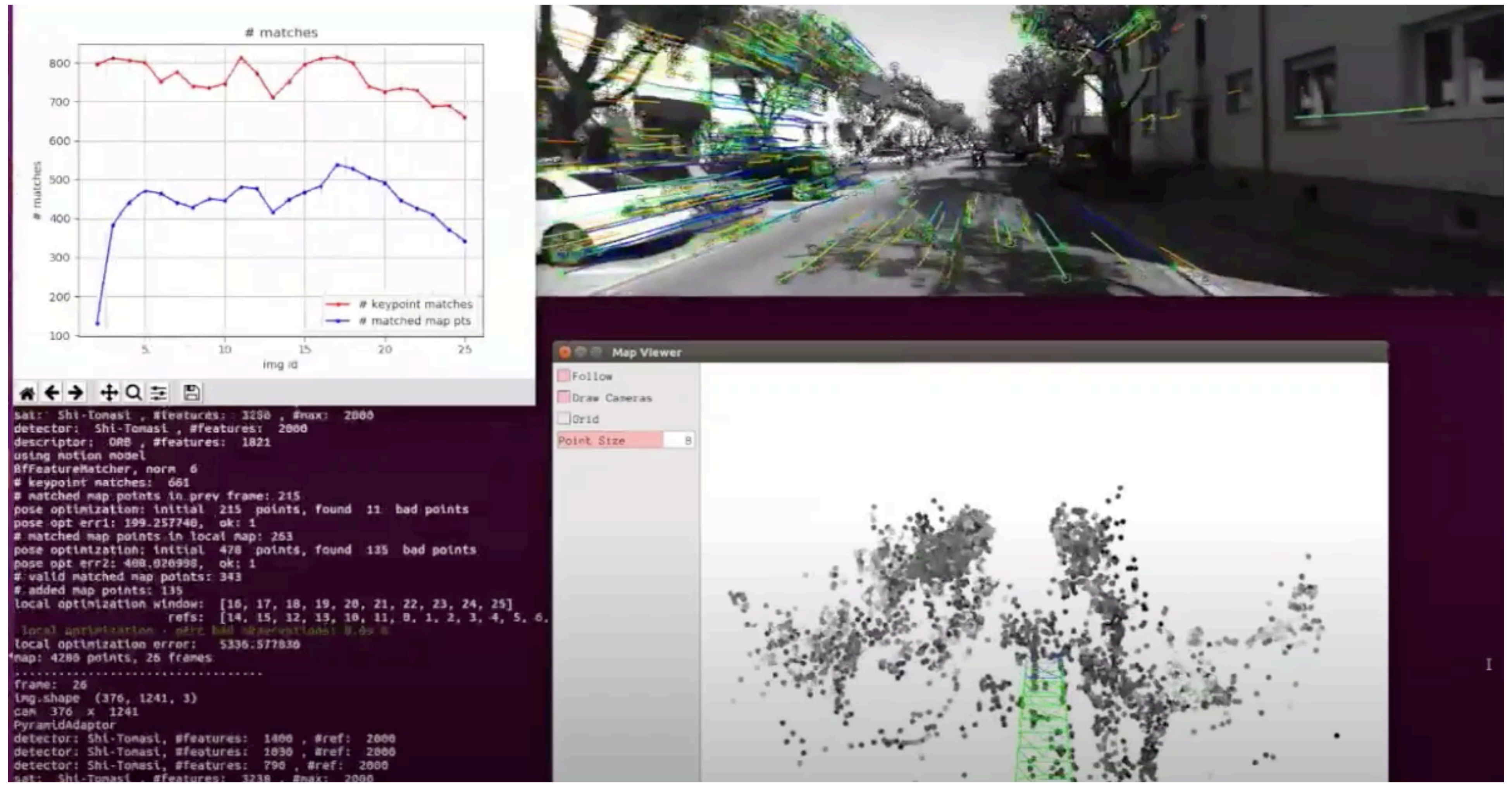
OrcVIO: Object residual constrained Visual-Inertial Odometry

ERL实验室

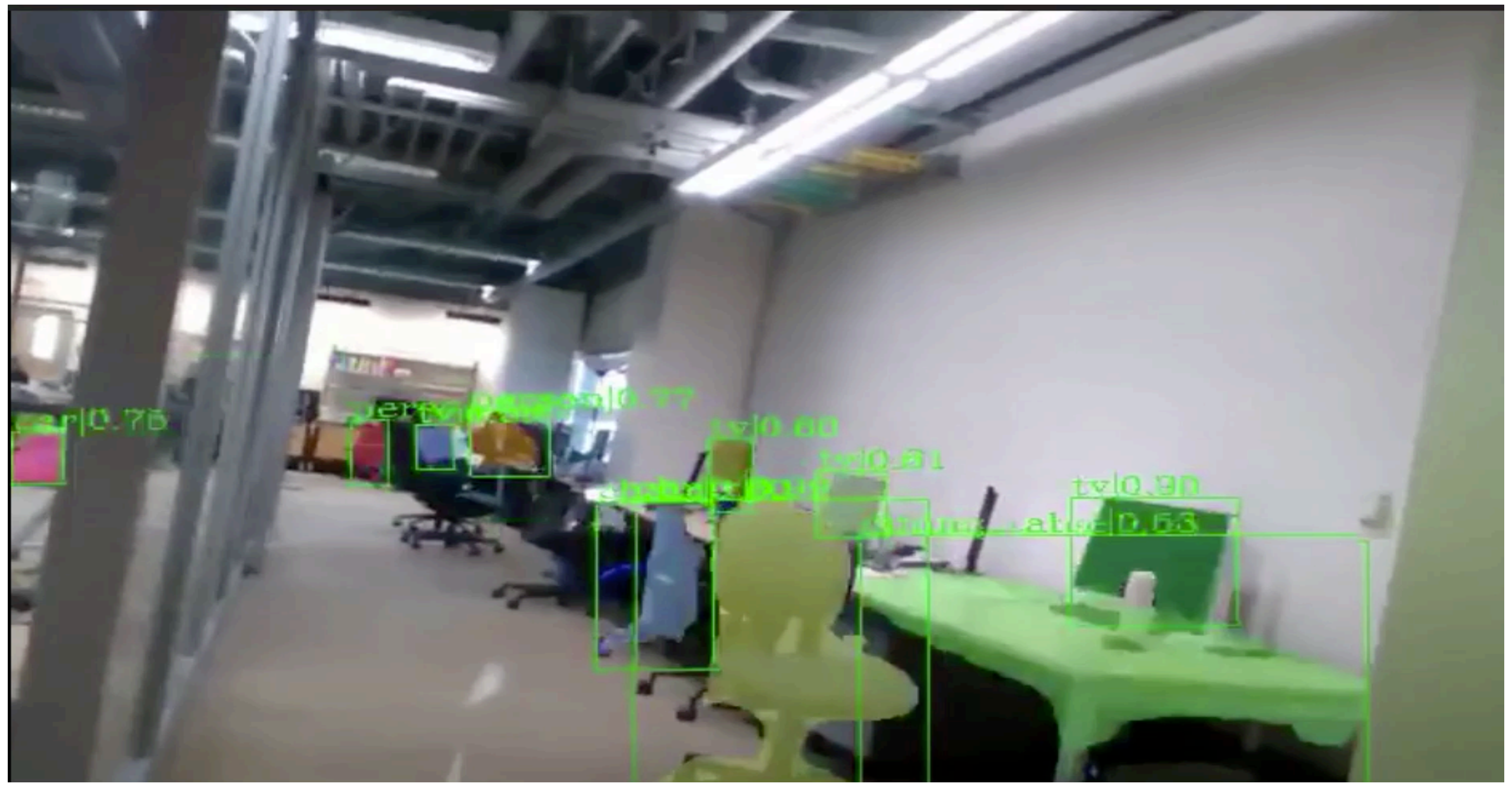
加州大学圣地亚哥分校

研究背景

- 主流的SLAM和视觉惯性里程计可以提供精确的环境几何信息



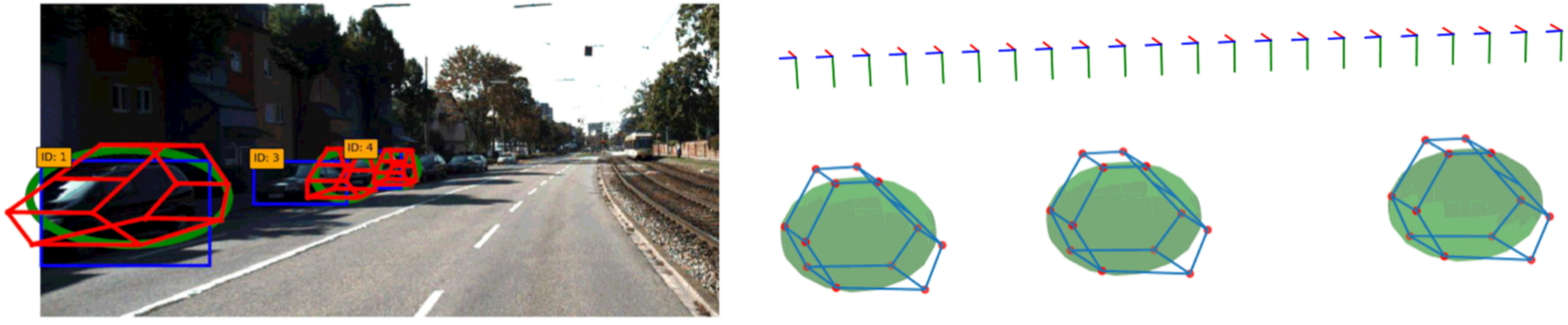
- 深度神经网络有强大的物体识别能力



研究背景

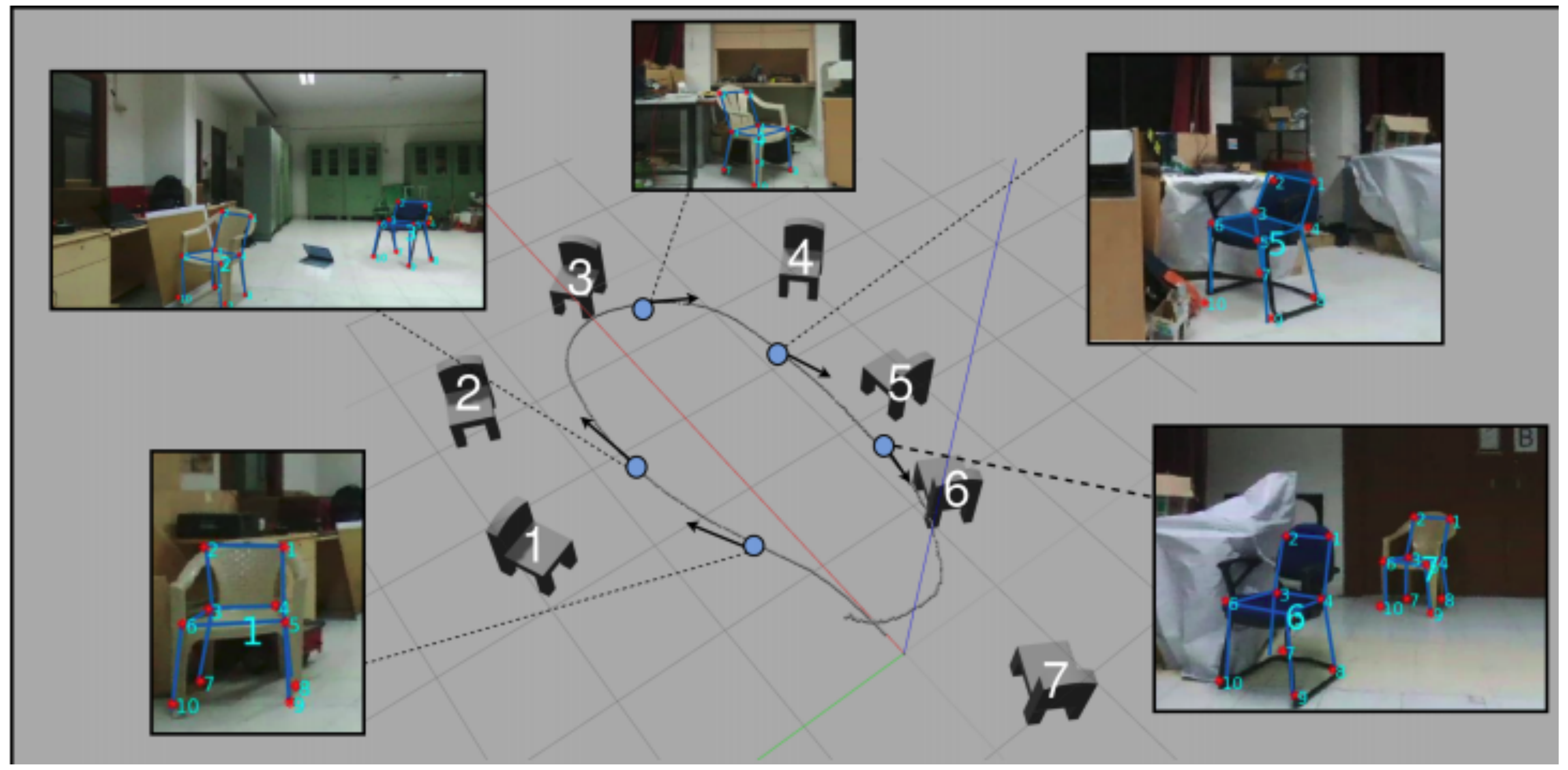
- 融合视觉惯性里程计的精确定位和建图，和深度学习得到的语义信息
- 我们提出了OrcVIO，运用物体残差帮助视觉惯性里程计定位和建图
- OrcVIO的输入是图像和惯性测量，输出是精确的定位和物体级别的语义地图

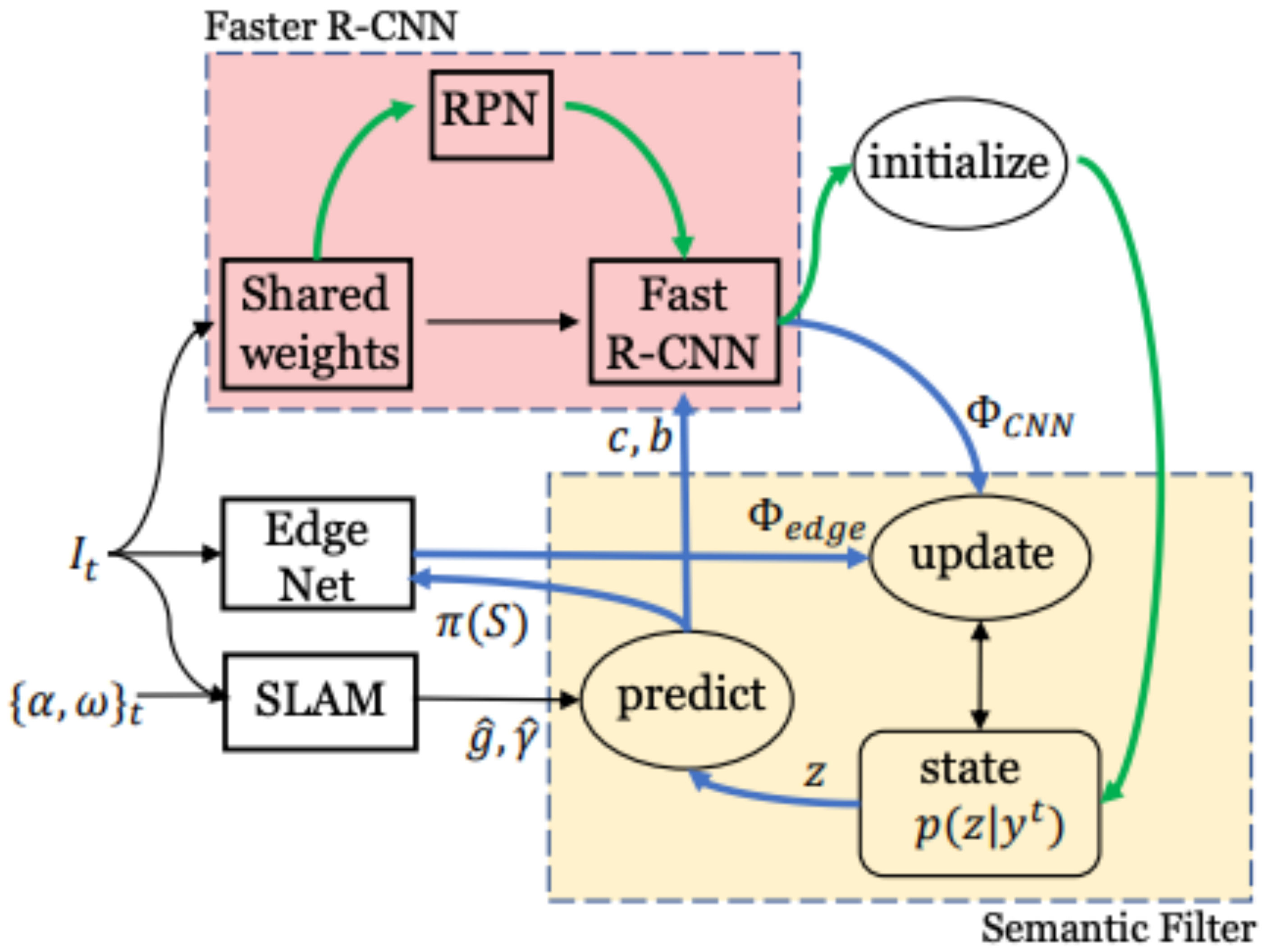
OrcVIO



相关工作

- 针对特定类别的方法，使用的是物体的特征点，和CAD模型



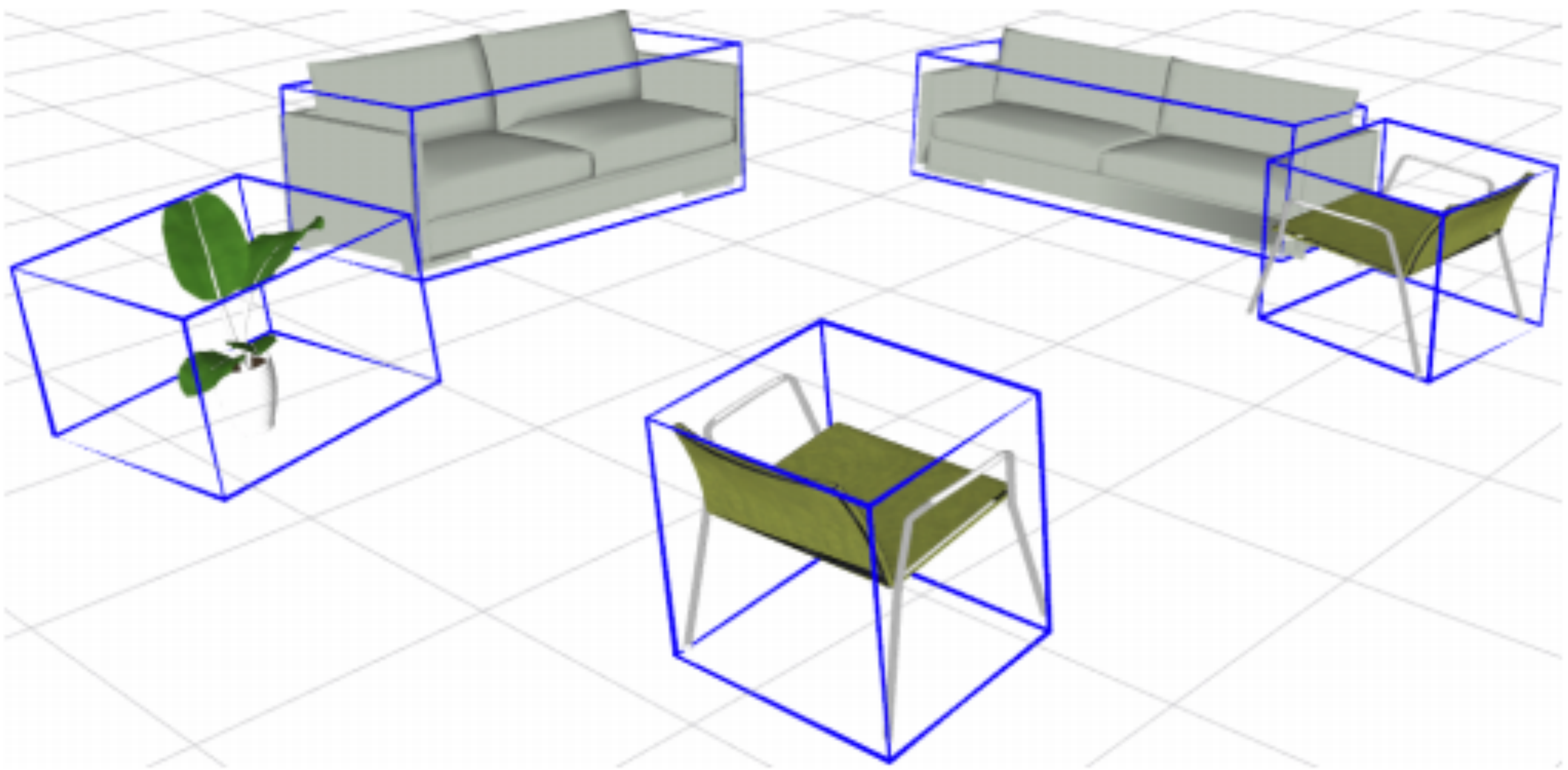


Parkhiya et al., 2018, ICRA　　Fei, X., & Soatto, S., 2018, ECCV

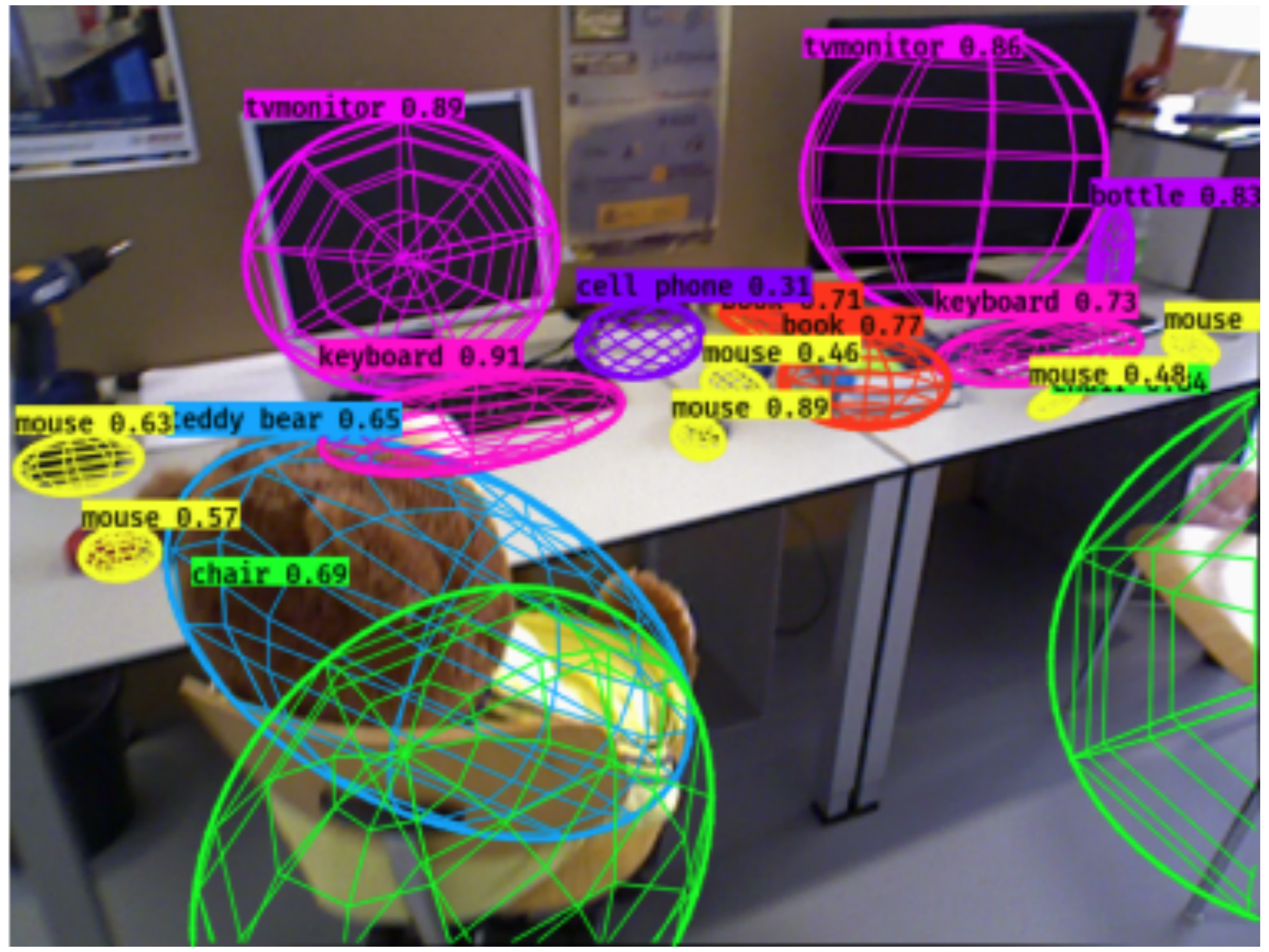
- Parkhiya, P., Khawad, R., Murthy, J.K., Bhowmick, B. and Krishna, K.M., 2018, May. Constructing category-specific models for monocular object-SLAM. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*
- Fei, X. and Soatto, S., 2018. Visual-inertial object detection and mapping. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 301-317).

相关方法

- 不针对类别的通用方法，使用的是几何模型，比如立方体和椭球



CubeSLAM, Yang, S. and Scherer, S., 2019, TRO

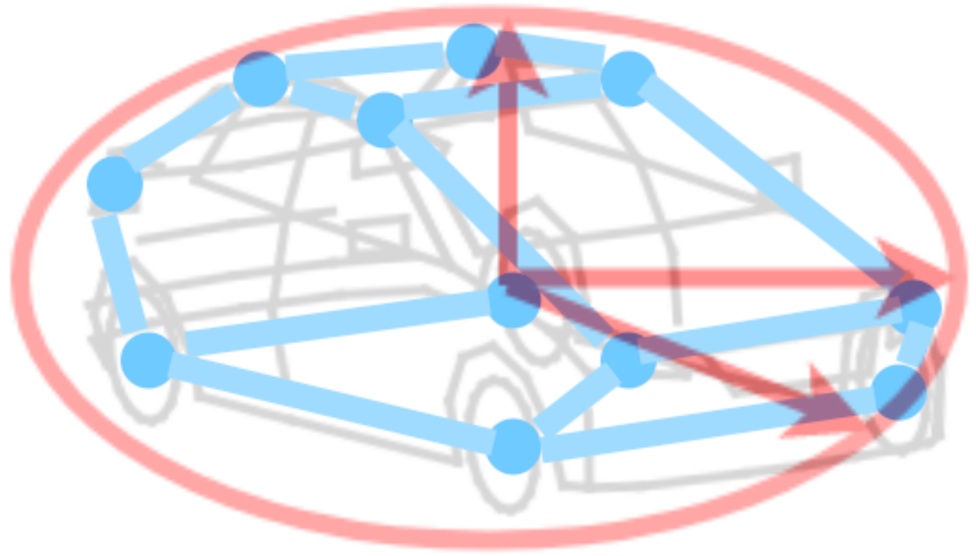


QuadricSLAM, Nicholson et al., 2018, RAL

- Yang, S. and Scherer, S., 2019. Cubeslam: Monocular 3-d object slam. IEEE Transactions on Robotics, 35(4), pp.925-938.
- Nicholson, L., Milford, M. and Sünderhauf, N., 2018. Quadricslam: Dual quadrics from object detections as landmarks in object-oriented slam. IEEE Robotics and Automation Letters, 4(1), pp.1-8.

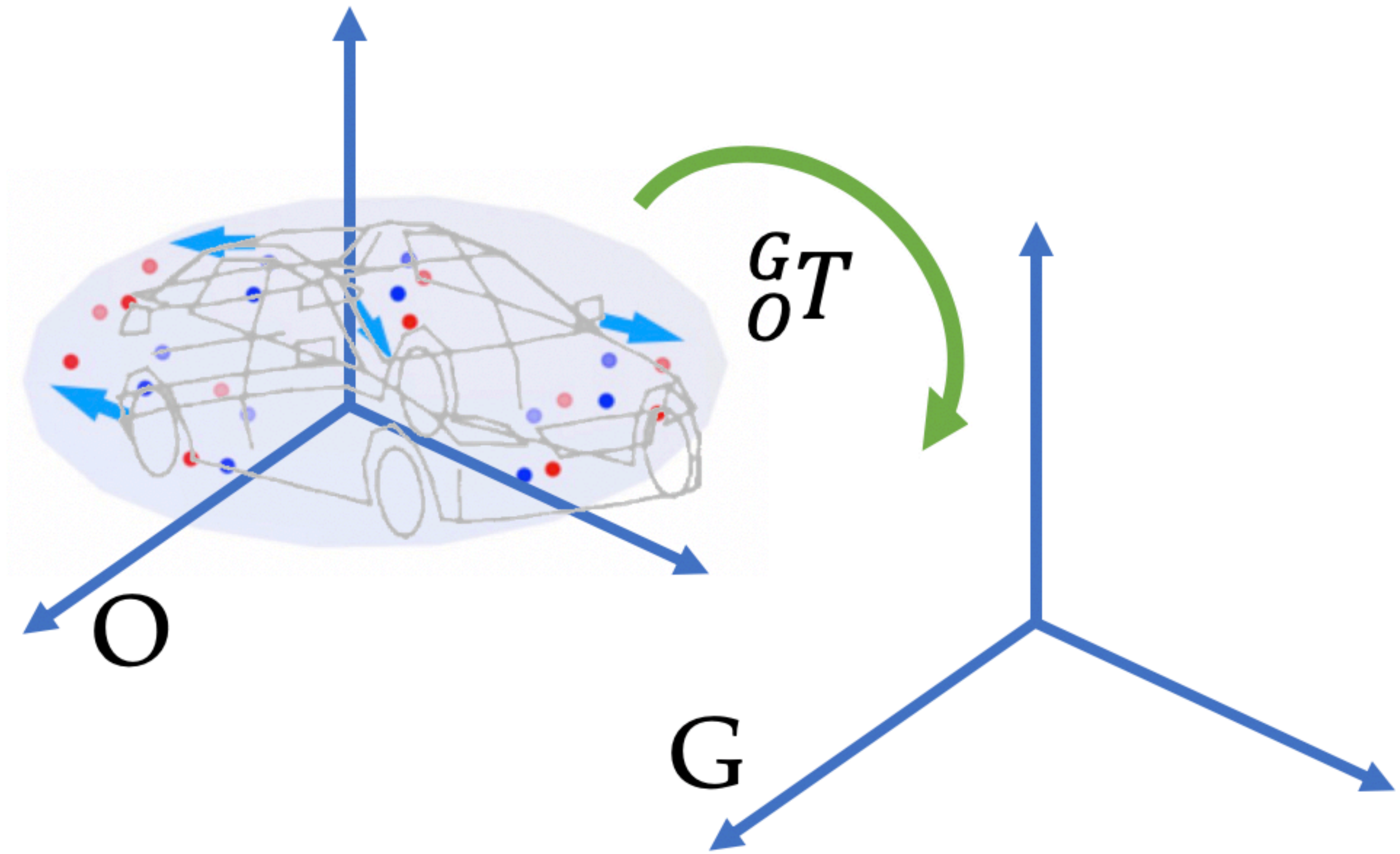
物体类别表示

- 粗略的物体表示：椭球（红色）
- 精细的物体表示：语义特征点（蓝色）



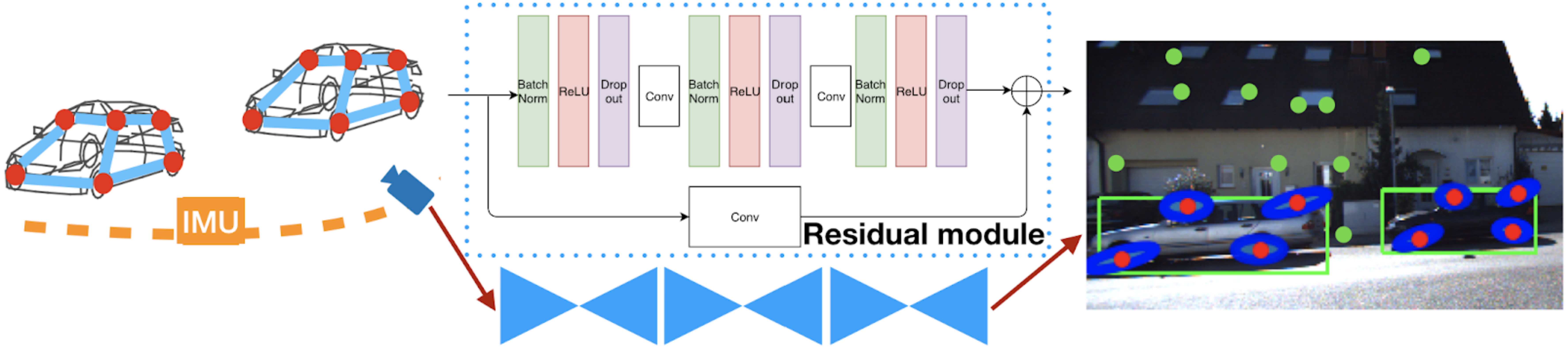
物体实例表示

- 语义特征点和椭球形状的变化（蓝色箭头）
- 物体的位姿 (绿色箭头)



问题描述

- 使用惯性测量，几何特征点，语义特征点，物体检测框，计算机器人位姿，构建几何地图，和物体级别的语义地图



(a) 3D scene (b) Stacked hourglass network (c) Visual observation

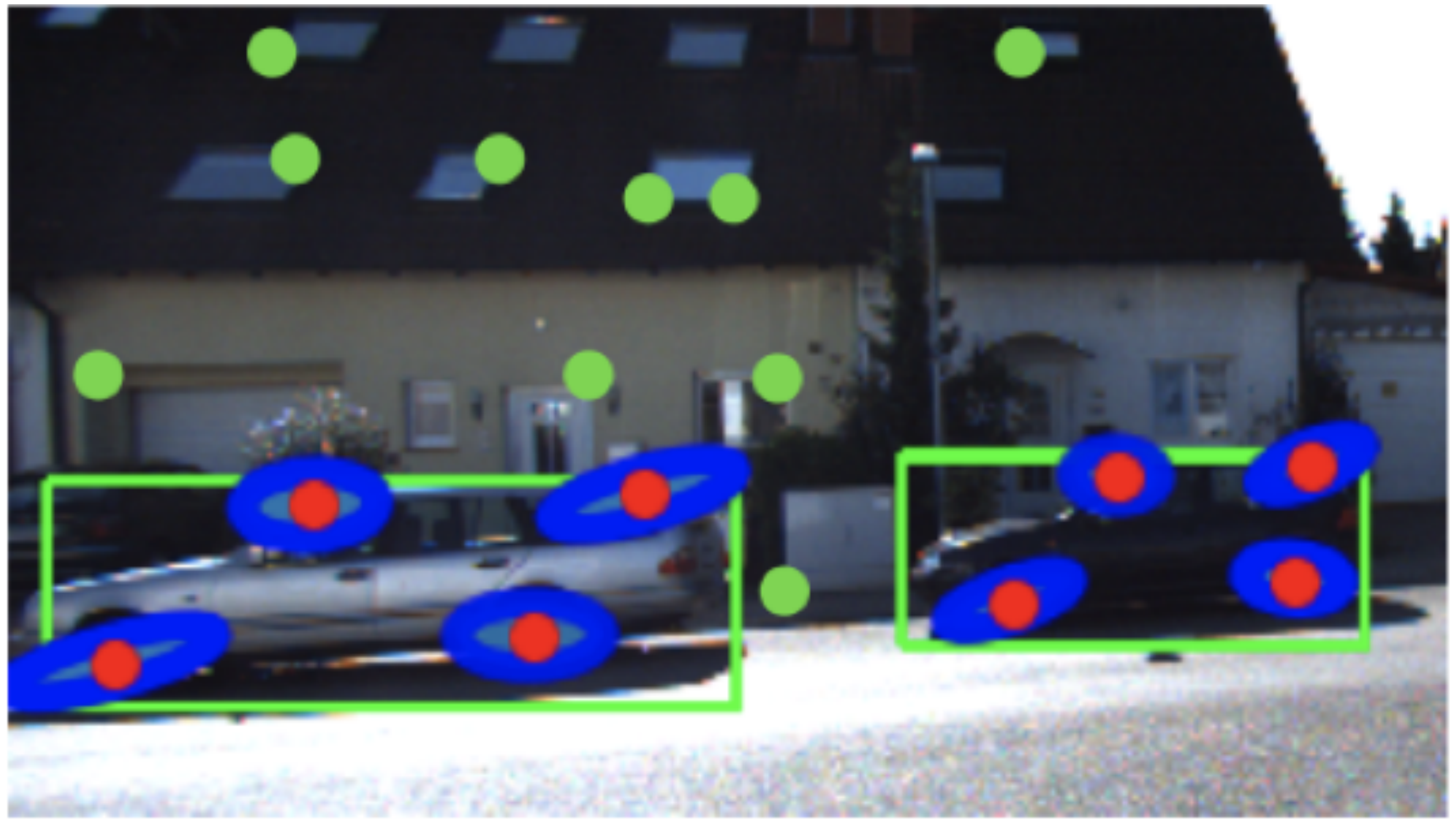
min TrajectoryCost + GeometricReprojectionCost + SemanticReprojectionCost + BoundingBoxCost + ShapeRegularization

目标函数

Problem. Determine the sensor trajectory $\mathcal{X}^*$, geometric landmarks $\mathcal{L}^*$, and object states $\mathcal{O}^*$ that minimize the weighted sum of squared errors:

$$\min_{\mathcal{X},\mathcal{L},\mathcal{O}} {}^{i}w \sum_{t} \|{}^{i}\mathbf{e}_{t,t+1}\|^2_{{}^{i}\mathbf{V}} + {}^{g}w \sum_{t,m,n} \mathbb{1}_{t,m,n} \|{}^{g}\mathbf{e}_{t,m,n}\|^2_{{}^{g}\mathbf{V}}$$
$$+ {}^{s}w \sum_{t,i,j,k} \mathbb{1}_{t,i,k} \|{}^{s}\mathbf{e}_{t,i,j,k}\|^2_{{}^{s}\mathbf{V}} + {}^{b}w \sum_{t,i,j,k} \mathbb{1}_{t,i,k} \|{}^{b}\mathbf{e}_{t,i,j,k}\|^2_{{}^{b}\mathbf{V}}$$
$$+ {}^{r}w \sum_{i} \|{}^{r}\mathbf{e}(\mathbf{o}_i)\|^2$$

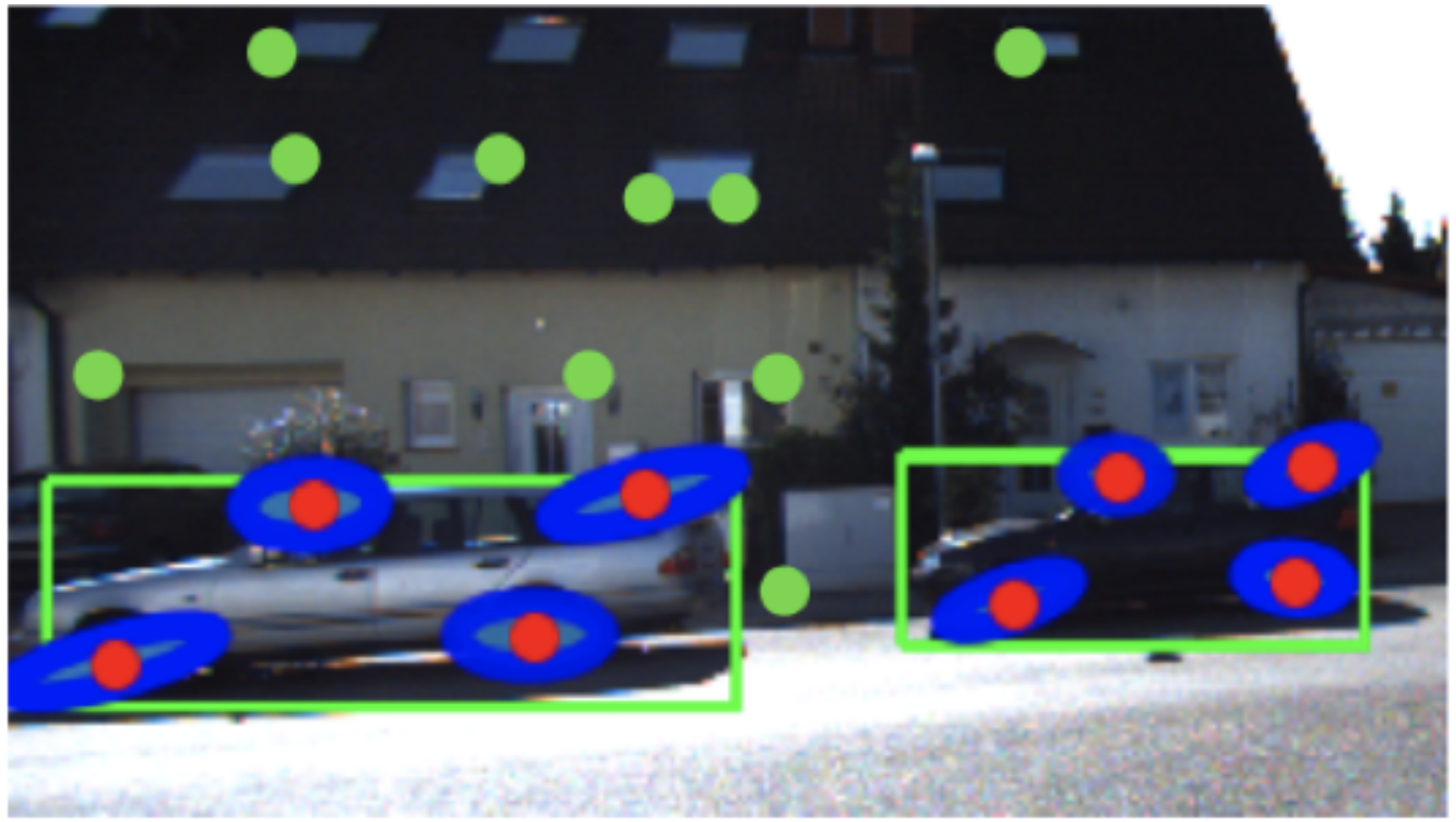
几何特征点



Define the geometric keypoint error as the difference between the image projection of a geometric landmark $\boldsymbol{\ell}$ using camera pose ${}_C\mathbf{T}$ and its associated keypoint observation ${}^g\mathbf{z}$:

$$^g\mathbf{e}\left(\mathbf{x}, \boldsymbol{\ell}, {}^g\mathbf{z}\right) \triangleq \mathbf{P}\pi\left({}_C\mathbf{T}^{-1}\underline{\boldsymbol{\ell}}\right) - {}^g\mathbf{z},$$

语义特征点

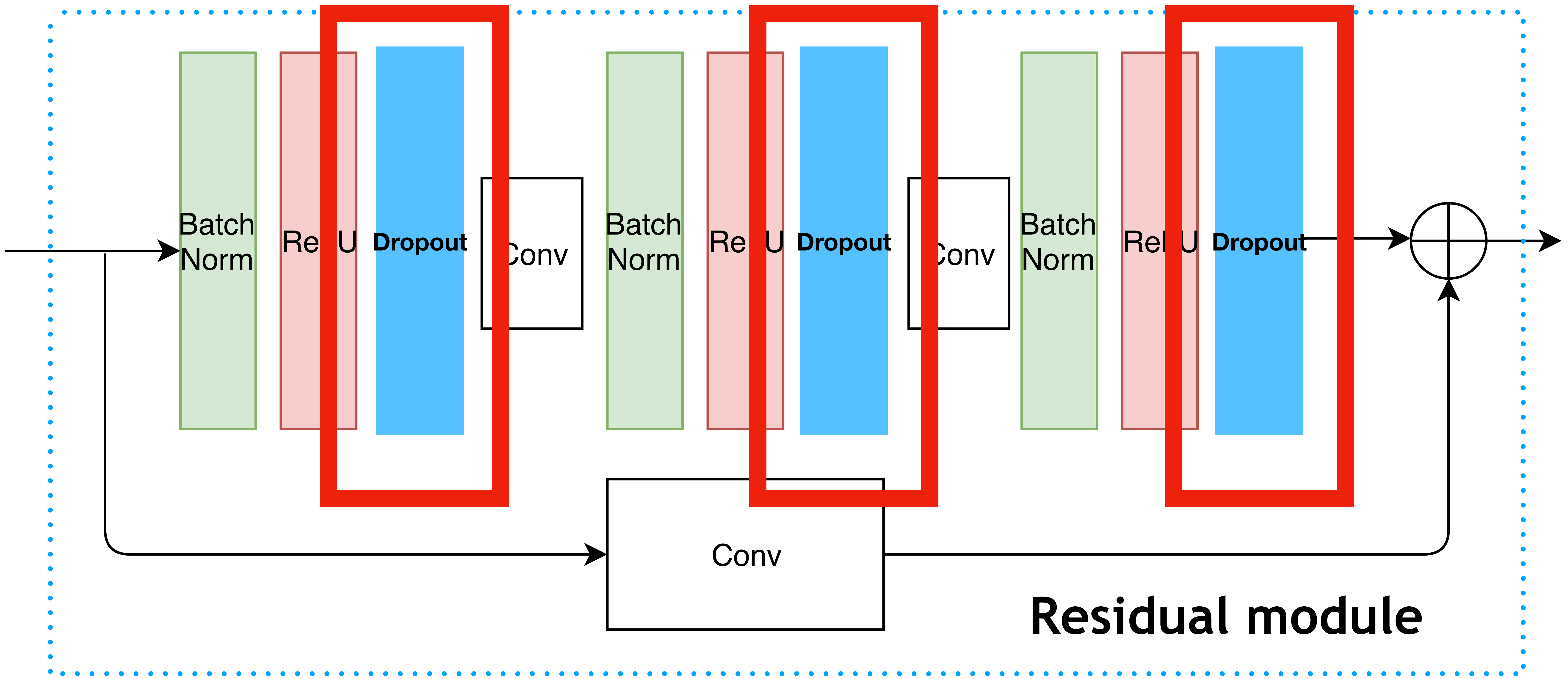


The semantic-keypoint error is defined as the difference between a semantic landmark $\mathbf{s}_j + \delta\mathbf{s}_j$, projected to the image plane using instance pose ${}_O\mathbf{T}$ and camera pose ${}_C\mathbf{T}_t$, and its corresponding semantic keypoint observation ${}^s\mathbf{z}_{t,j,k}$:

$$ {}^s\mathbf{e}(\mathbf{x}_t, \mathbf{o}, {}^s\mathbf{z}_{t,j,k}) \triangleq \mathbf{P}\pi\left({}_C\mathbf{T}_t^{-1}\,{}_O\mathbf{T}\left(\underline{\mathbf{s}}_j + \delta\underline{\mathbf{s}}_j\right)\right) - {}^s\mathbf{z}_{t,j,k}. $$

语义特征点

- 使用StarMap深度神经网络来检测语义特征点
- 添加了Dropout层来获取特征点位置的协方差



- Zhou, X., Karpur, A., Luo, L. and Huang, Q., 2018. Starmap for category-agnostic keypoint and viewpoint estimation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 318-334).

语义特征点

- 使用卡尔曼滤波在物体级别跟踪语义特征点

物体状态的初始化

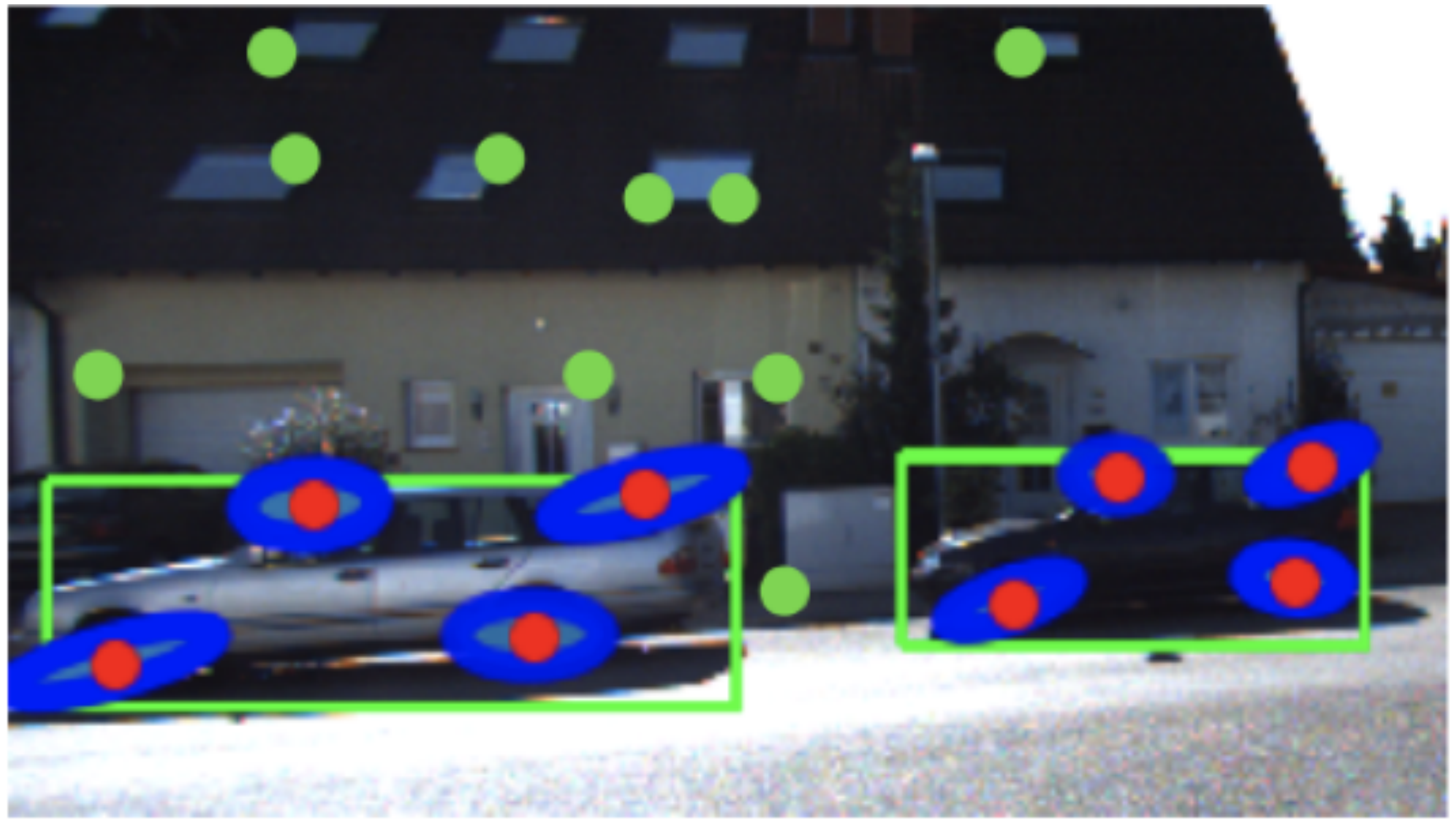
$$\mathbf{0} = \mathbf{P}_C \hat{\mathbf{T}}_t^{-1}{}_O \hat{\mathbf{T}} \underline{\mathbf{s}}_j - \lambda_{t,j,k}{}^s\mathbf{z}_{t,j,k}$$

Rearranging that leads to

$$\begin{aligned} {}_C\hat{\mathbf{R}}_t^\top \left(\boldsymbol{\xi}_j - {}_C\hat{\mathbf{p}}_t\right) &= \lambda_{t,j,k}{}^s\mathbf{z}_{t,j,k} \\ {}_C\hat{\mathbf{R}}_t^\top \boldsymbol{\xi}_j - {}^s\mathbf{z}_{t,j,k}\lambda_{t,j,k} &= {}_C\hat{\mathbf{R}}_t^\top {}_C\hat{\mathbf{p}}_t \\ \boldsymbol{\xi}_j - {}_C\hat{\mathbf{R}}_t{}^s\mathbf{z}_{t,j,k}\lambda_{t,j,k} &= {}_C\hat{\mathbf{p}}_t \end{aligned}$$

Tracked Targets

物体检测框



To define a bounding-box error, we observe that if the dual ellipsoid $\mathbf{Q}^*_{(\mathbf{u}+\delta\mathbf{u})}$ of instance $\mathbf{i}$ is estimated accurately, then the lines ${}^b\underline{\mathbf{z}}_{t,j,k}$ of the k-th bounding-box at time t should be tangent to the image plane conic projection of $\mathbf{Q}^*_{(\mathbf{u}+\delta\mathbf{u})}$:

$${}^b\mathbf{e}(\mathbf{x},\mathbf{o},{}^b\underline{\mathbf{z}}) \triangleq {}^b\underline{\mathbf{z}}^\top \mathbf{P}\,{}_C\mathbf{T}^{-1}\,{}_O\mathbf{T}\mathbf{Q}^*_{(\mathbf{u}+\delta\mathbf{u})}\,{}_O\mathbf{T}^\top\,{}_C\mathbf{T}^{-\top}\mathbf{P}^\top\,{}^b\underline{\mathbf{z}}.$$

雅克比矩阵

$$\frac{\partial^s \mathbf{e}}{\partial_O \boldsymbol{\xi}} = \mathbf{P}\frac{d\pi}{d\underline{\mathbf{s}}}\left({}_C\hat{\mathbf{T}}_t^{-1}{}_O\hat{\mathbf{T}}\left(\underline{\mathbf{s}}_j + \underline{\delta\hat{\mathbf{s}}}_j\right)\right){}_C\hat{\mathbf{T}}_t^{-1}\left[{}_O\hat{\mathbf{T}}\left(\underline{\mathbf{s}}_j + \underline{\delta\hat{\mathbf{s}}}_j\right)\right]^{\odot}$$

$$\frac{\partial^s \mathbf{e}}{\partial\delta\tilde{\mathbf{s}}_j} = \mathbf{P}\frac{d\pi}{d\underline{\mathbf{s}}}\left({}_C\hat{\mathbf{T}}_t^{-1}{}_O\hat{\mathbf{T}}\left(\underline{\mathbf{s}}_j + \underline{\delta\hat{\mathbf{s}}}_j\right)\right){}_C\hat{\mathbf{T}}_t^{-1}{}_O\hat{\mathbf{T}}\begin{bmatrix}\mathbf{I}_3\\ \mathbf{0}^\top\end{bmatrix} \in \mathbb{R}^{2\times 3}.$$

$$\frac{\partial^b \mathbf{e}}{\partial_O \boldsymbol{\xi}} = 2^b\underline{\mathbf{z}}^\top \mathbf{P}{}_C\hat{\mathbf{T}}_t^{-1}{}_O\hat{\mathbf{T}}\hat{\mathbf{Q}}^*_{(\mathbf{u}+\delta\hat{\mathbf{u}})}{}_O\hat{\mathbf{T}}^\top\left[{}_C\hat{\mathbf{T}}_t^{-\top}\mathbf{P}^{\top b}\underline{\mathbf{z}}\right]^{\circledcirc}$$

$$\frac{\partial^b \mathbf{e}}{\partial\delta\tilde{\mathbf{u}}} = \left(2(\mathbf{u}+\delta\hat{\mathbf{u}})\odot\mathbf{y}\odot\mathbf{y}\right)^\top \in \mathbb{R}^{1\times 3}$$

$$\mathbf{y} \triangleq \begin{bmatrix}\mathbf{I}_3 & \mathbf{0}\end{bmatrix}{}_O\hat{\mathbf{T}}^\top{}_C\hat{\mathbf{T}}_t^{-\top}\mathbf{P}^{\top b}\underline{\mathbf{z}}.$$

视觉惯性里程计

- 使用基于MSCKF的框架融合视觉和惯性测量信息，估计机器人姿态，构建几何地图
- 不再使用四元数，而是直接使用旋转矩阵来表示位姿
- 关于机器人的运动方程，我们不使用数值近似，而是提出了完整的表达形式和推导

$$ {}_I\mathbf{x}_t \triangleq \left({}_I\mathbf{R}_t,\ {}_I\mathbf{p}_t,\ {}_I\mathbf{v}_t,\ \mathbf{b}_g,\ \mathbf{b}_a\right) $$

$$ {}_I\hat{\mathbf{p}}^p_{t+1} = {}_I\hat{\mathbf{p}}_t + {}_I\hat{\mathbf{v}}_t\tau + \mathbf{g}\frac{\tau^2}{2} + {}_I\hat{\mathbf{R}}_t\mathbf{H}_L\left(\tau\left({}^i\boldsymbol{\omega}_t - \hat{\mathbf{b}}_{g,t}\right)\right)\left({}^i\mathbf{a}_t - \hat{\mathbf{b}}_{a,t}\right)\tau^2 $$
$$ {}_I\hat{\mathbf{v}}^p_{t+1} = {}_I\hat{\mathbf{v}}_t + \mathbf{g}\tau + {}_I\hat{\mathbf{R}}_t\mathbf{J}_L\left(\tau\left({}^i\boldsymbol{\omega}_t - \hat{\mathbf{b}}_{g,t}\right)\right)\left({}^i\mathbf{a}_t - \hat{\mathbf{b}}_{a,t}\right)\tau $$

$$ \mathbf{J}_L(\boldsymbol{\omega}) = \mathbf{I}_3 + \frac{1-\cos\|\boldsymbol{\omega}\|}{\|\boldsymbol{\omega}\|^2}\boldsymbol{\omega}_\times + \frac{\|\boldsymbol{\omega}\| - \sin\|\boldsymbol{\omega}\|}{\|\boldsymbol{\omega}\|^3}\boldsymbol{\omega}_\times^2 $$
$$ \mathbf{H}_L(\boldsymbol{\omega}) = \frac{1}{2}\mathbf{I}_3 + \frac{\|\boldsymbol{\omega}\| - \sin\|\boldsymbol{\omega}\|}{\|\boldsymbol{\omega}\|^3}\boldsymbol{\omega}_\times + \frac{2(\cos\|\boldsymbol{\omega}\| - 1) + \|\boldsymbol{\omega}\|^2}{2\|\boldsymbol{\omega}\|^4}\boldsymbol{\omega}_\times^2. $$

结果展示

- 几何特征点（绿色），物体检测框和语义特征点（彩色）的跟踪

结果展示

- 物体地图的二维投影（语义特征点为红色，椭球为绿色）

定量分析

TABLE II

PRECISION-RECALL EVALUATION ON KITTI OBJECT SEQUENCES

	Translation error →	≤ 0.5 m		≤ 1.0 m		≤ 1.5 m	
Rotation error	Method	Precision	Recall	Precision	Recall	Precision	Recall
$\leq 30°$	SubCNN [36]	0.10	0.07	0.26	0.17	0.38	0.26
	VIS-FNL [14]	**0.14**	0.10	**0.34**	**0.24**	**0.49**	**0.35**
	OrcVIO	0.10	**0.12**	0.18	0.21	0.22	0.25
$\leq 45°$	SubCNN [36]	0.10	0.07	0.26	0.17	0.38	0.26
	VIS-FNL [14]	**0.15**	0.11	**0.35**	0.25	**0.50**	**0.36**
	OrcVIO	**0.15**	**0.17**	0.25	**0.28**	0.31	0.35
–	SubCNN [36]	0.10	0.07	0.27	0.18	0.41	0.28
	VIS-FNL [14]	0.16	0.11	0.40	0.29	0.58	0.42
	OrcVIO	**0.29**	**0.33**	**0.50**	**0.56**	**0.62**	**0.69**

谢谢观看！

项目主页：http://me-llamo-sean.cf/orcvio_githubpage/

OrcVIO

ERL

CONTEXTUAL ROBOTICS INSTITUTE

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering