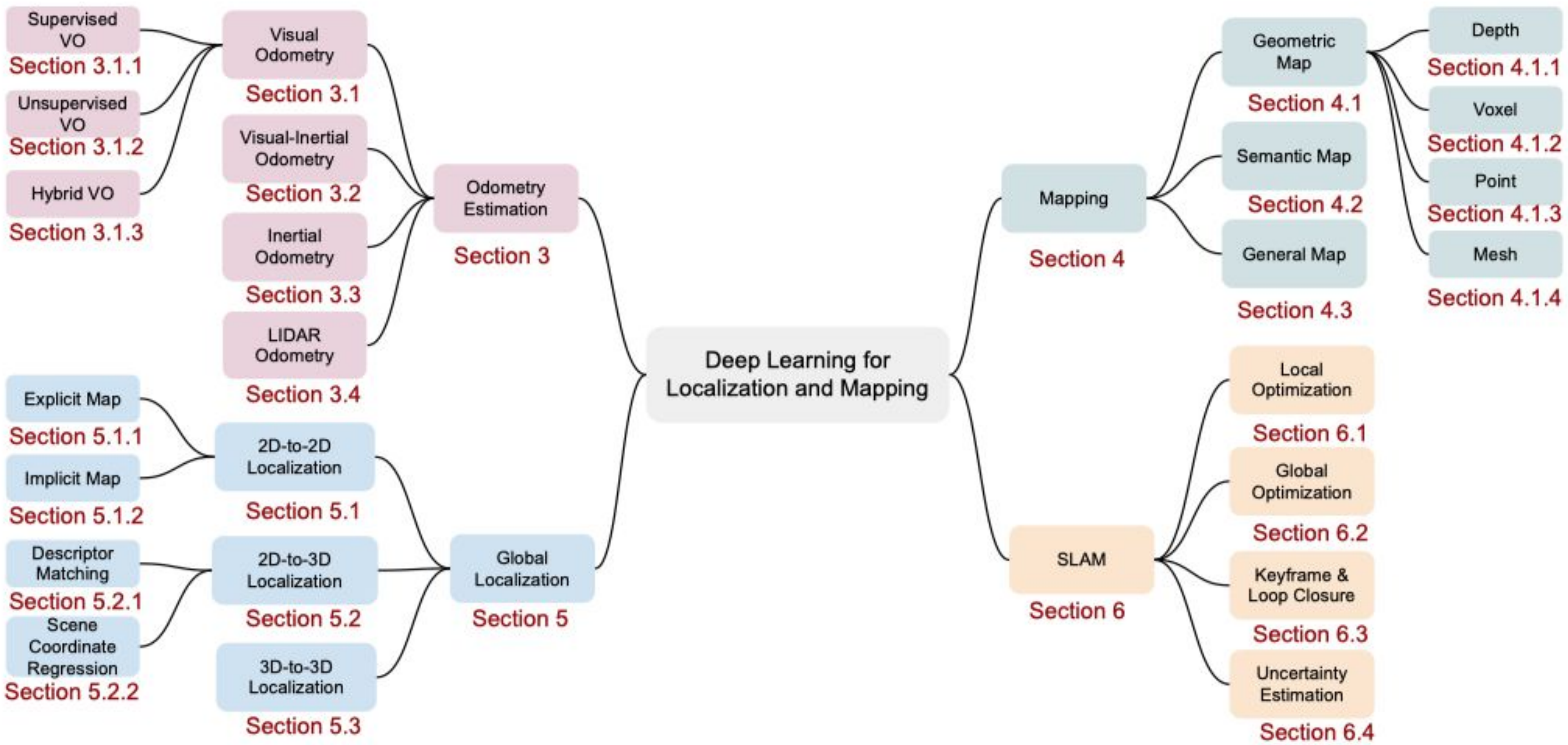
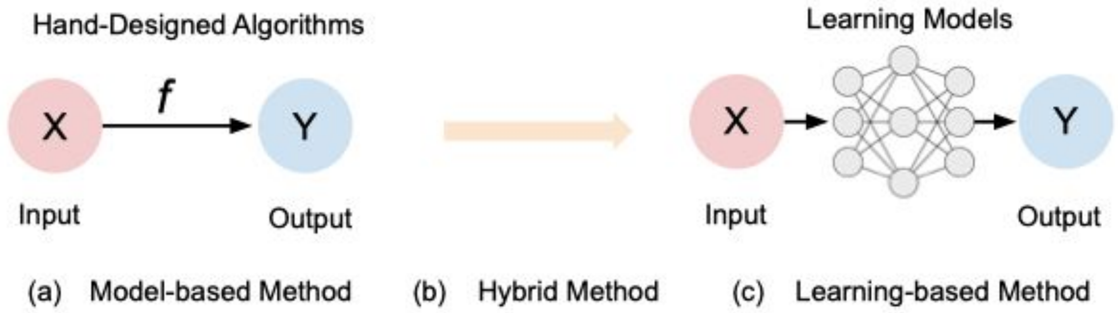


Figure 2: A taxonomy of existing works on deep learning for localization and mapping.



X: sensor data, e.g. images, inertial data, LIDAR point clouds,
Y: target values, e.g. relative motion (translation & rotation), global pose (location & orientation), scene geometry & semantics

Why to Study Deep Learning for Localization and Mapping?

- Imperfect sensor measurements, inaccurate system modelling, complex environmental dynamics and unrealistic constraints impact both the accuracy and reliability of hand-crated systems.

- Learning methods can leverage highly expressive deep neural network as an universal approximator, and automatically discover features relevant to task.

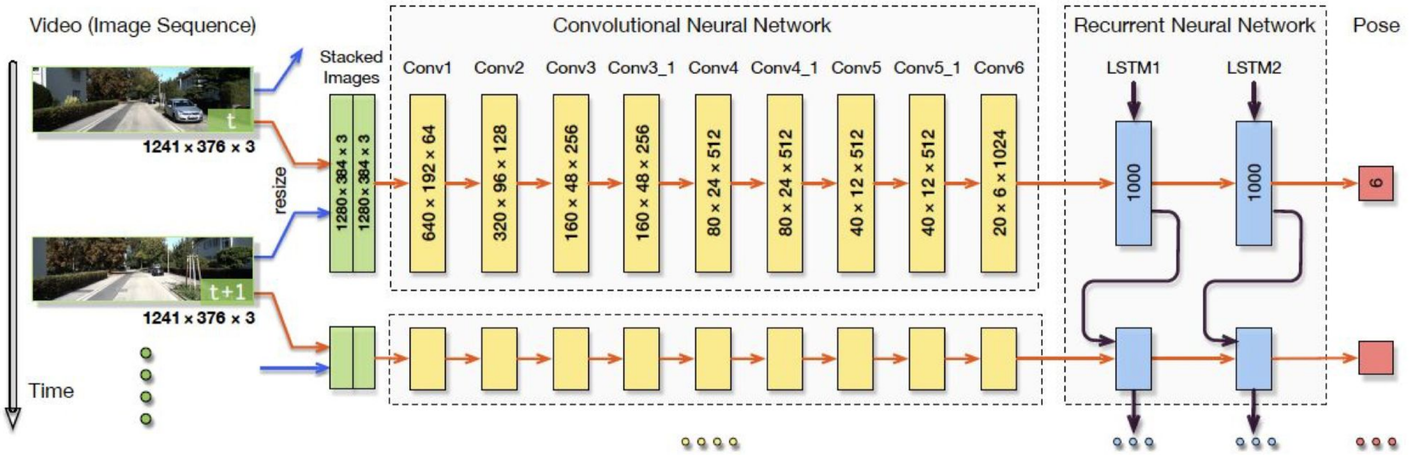
- Learning methods allow systems to learn from past experience, and actively exploit new information.

- Its capability of fully exploiting the increasing amount of sensor data and computational power.

Odometry estimation

- end-to-end VO
  - Supervised
  - Unsupervised
- hybrid VO

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \|\hat{\mathbf{p}}_t - \mathbf{p}_t\|_2^2 + \|\hat{\boldsymbol{\varphi}}_t - \boldsymbol{\varphi}_t\|_2^2,$$



(a) Supervised learning of visual odometry

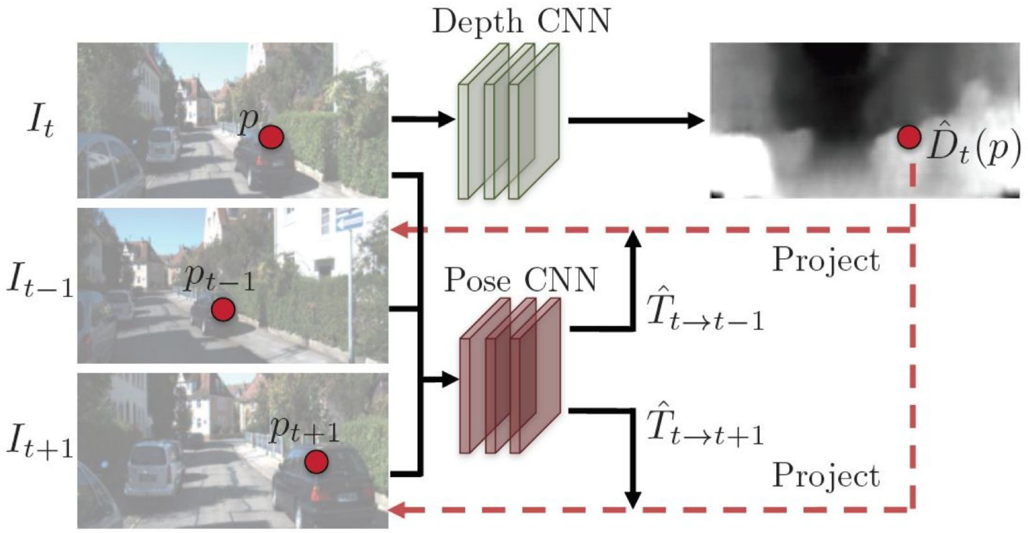
Preserves scale!

Odometry estimation

- end-to-end VO
  - Supervised
  - Unsupervised
- hybrid VO

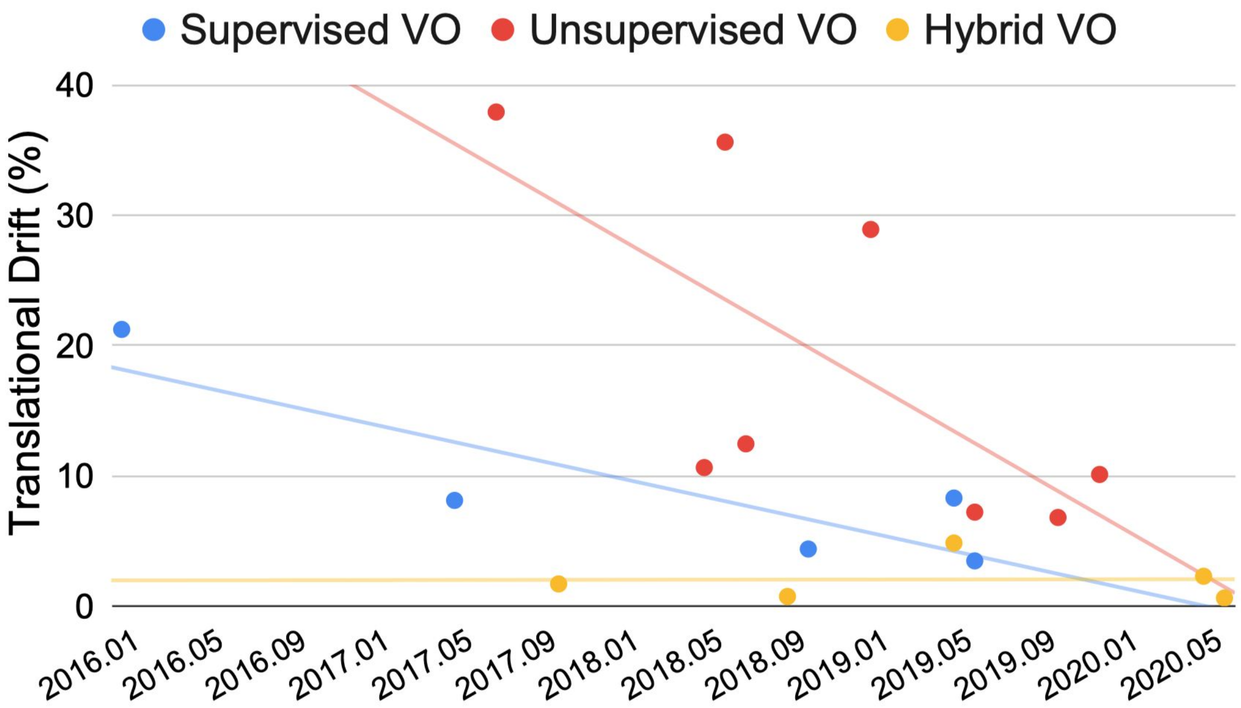
$$p_s \sim \mathbf{K}\mathbf{T}_{t\rightarrow s}\mathbf{D}_t(p_t)\mathbf{K}^{-1}p_t$$

$$\mathcal{L}_{\text{photo}} = \sum_{<\mathbf{I}_1,\ldots,\mathbf{I}_N>\in S} \sum_p |\mathbf{I}_t(p) - \hat{\mathbf{I}}_s(p)|,$$



(b) Unsupervised learning of visual odometry

Odometry estimation



Although unsupervised VO still cannot compete with supervised VO in performance, as illustrated in Figure 5, its concerns of scale metric and scene dynamics problem have been largely resolved.

Odometry estimation

- end-to-end VO
  - Supervised
  - Unsupervised
- hybrid VO

Mapping

B Mapping

- B.1 Geometric Mapping
- B.2 Semantic Mapping
- B.3 General Mapping

- Depth
- Voxel
- Point
- Mesh