



Vacant Parking Space Detection based on Task Consistency and Reinforcement Learning

Manh-Hung Nguyen

Faculty of Electrical Electronic Engineering, HCM-UTE, Vietnam hungnm@hcm.ute.edu.vn

Tzu-Yin Chao

Department of Computer Science, National Chiao Tung University, Taiwan chaoziyin@gmail.com

Ching-Chun Huang

Department of Computer Science, National Chiao Tung University, Taiwan chingchun@cs.nctu.edu.tw



Contributions

We proposed a novel framework that allows the system to train a target model (e.g., a vacant-space detector) via the <u>task consistency</u> with a source model (e.g., a car motion classifier).

- Unlike transfer learning, the source model and the target model in our framework are not restricted to deal with the same type of task.
- > The proposed framework is suitable for online learning, which is lable-free (<u>unsupervised rewards</u>).
- We test the method on a parking lot scenario and <u>corrupted rewards</u> are filtered out automatically

Algorithm

Algorithm 1 Task Consistency learning (TCL) 1: **procedure** TCL(Require: $M^s_{\phi}(x^s_i)$, Environment, $f^{s}(s_{t}), f^{t}(s_{t})$; Output: target network θ) 2 $D_{online} = \emptyset$ Random initialize θ 3: while collect training trajectories do 4 ⊳ Data collection 5: Query s_t from the environment $x_t^s = f^s(s_t), x_t^t = f^t(s_t)$ ▷ Select source and 6: target input $\hat{y}_t^s = M_\phi^s(x_t^s)$ ▷ Predict source decision 7 if $P(y_t^s) > \delta$ then 8: Extract $x_i^t = f_t(s_i) \mid i = 1 \sim T$ from the 9: segment trajectory around the frame t^{th} . end if 10 Add the tuple $(x_i^s, \hat{y}_i^s, x_i^t)_{i=1:T}$ into D_{online} 11. end while 12: while Training do 13: for each training trajectory do 14: Run the target / policy model $\pi_{\theta}(a_t \mid x_t)$ to 15: get a_t or y_t 16 Estimate $r(a_t)$ by equation (5) Estimate baseline value by equation (6) 17: 18: Estimate the gradient in equation (4) Update the policy $\theta \leftarrow \theta + \alpha \nabla_{\theta} J(\theta)$ 19: end for 20: end while 21: Repeat the data collection process in the 4th line 22: Repeat the training process in the 13th line 23: 24: end procedure

Reward design







The number of clean/noisy trajectories under different thresholds identified by the baseline motion classifier and robust motion classifier. (Evaluate on 1530 trajectories)

 \cdot Cl: Clean, No: Noisy, MC: Motion Classifier

δ Cl No No Cl No Cl No No<
0.6 950 142 1406 27
0.90-
0.7 666 82 1292 11
0.8 393 33 1090 0

- We proposed a task consistency framework, which enables the system to learn a target task from a source task in a reinforcement learning manner.
- The framework has two benefits:
 - The source model and target model are not restricted to deal with the same type of task.
 - By applying reinforcement learning approach with unsupervised rewards, our framework is label-free.
- The framwork is applied to learn a vacant space detector based on a motion classifier:
 - The reward design is capable of filtering out some easy corrupted rewards automatically.

T/3

21/3