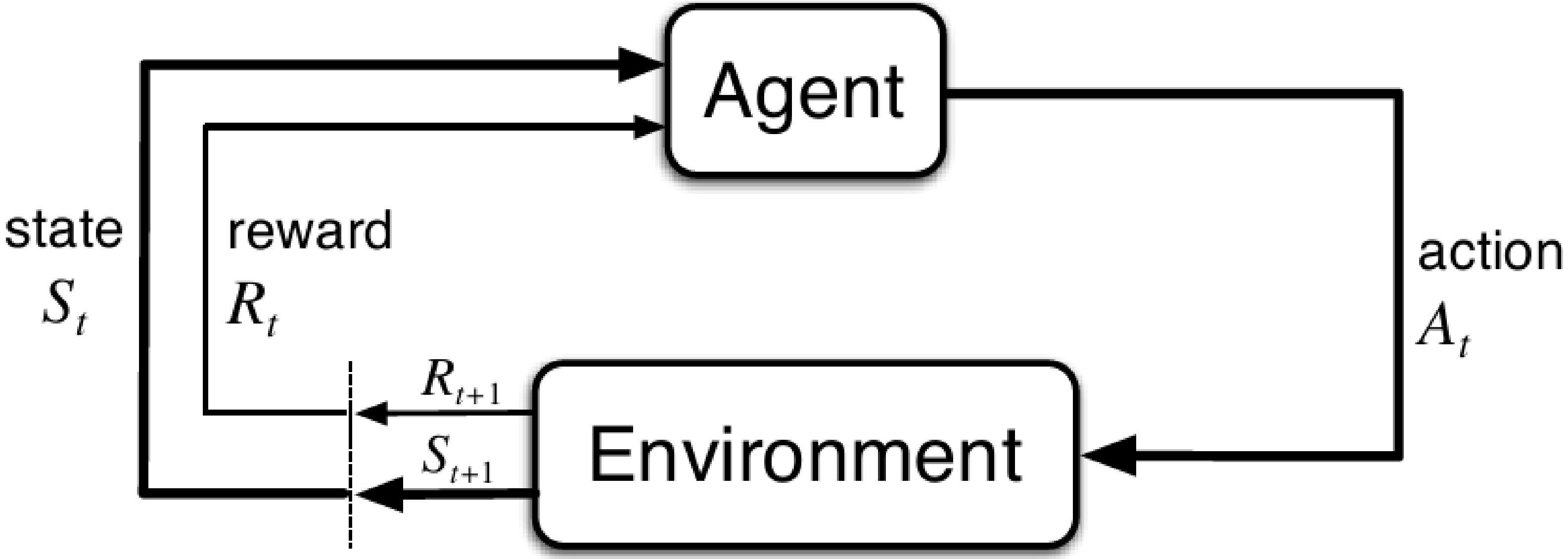
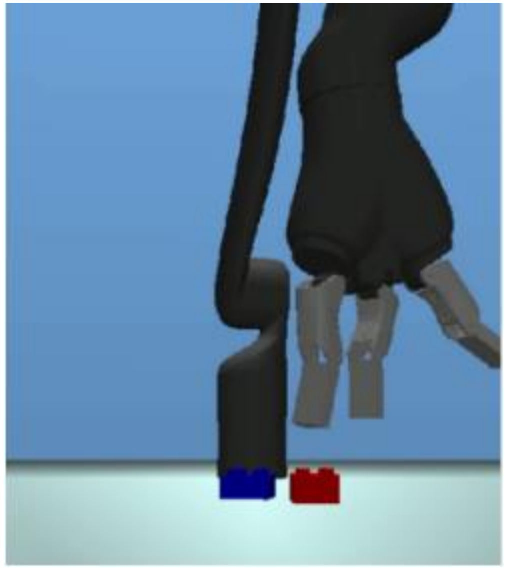
增强学习在导航中的应用

马留龙

增强学习



增强学习在机器人中的应用

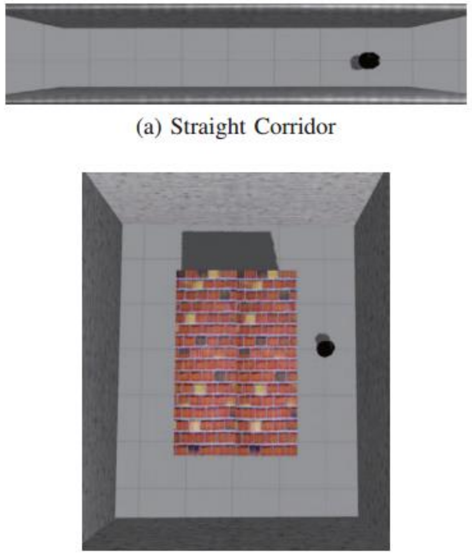


抓取

直升机控制

运动规划

A robot exploration strategy based on q-learning network



(a) Straight Corridor

(b) circular Corridor

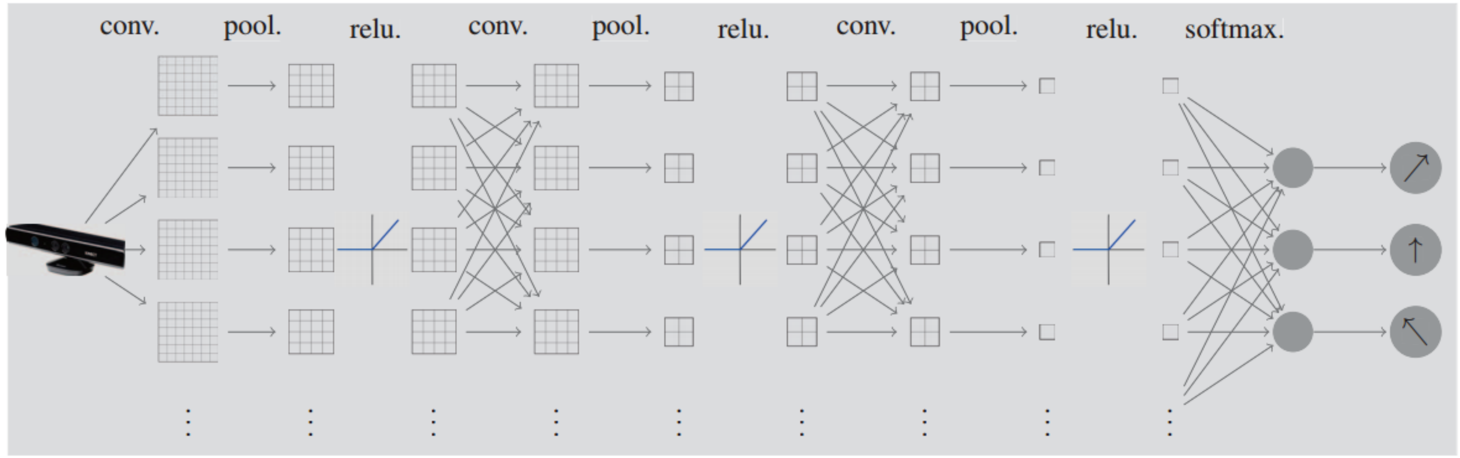


Fig. 1. Structure of the CNN layers. Depth images after down sampling will be fed into to the model. Three Convolution layers with pooling and rectifier layers after are connected together. After that, feature maps of every input will be fully connected and fed to the softmax layer of the classier.

Lei T, Ming L. A robot exploration strategy based on q-learning network[C]//Real-time Computing and Robotics (RCAR), IEEE International Conference on. IEEE, 2016: 57-62.

实现细节

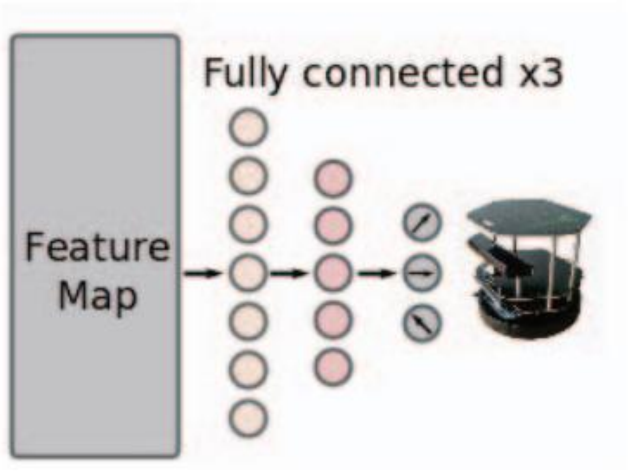
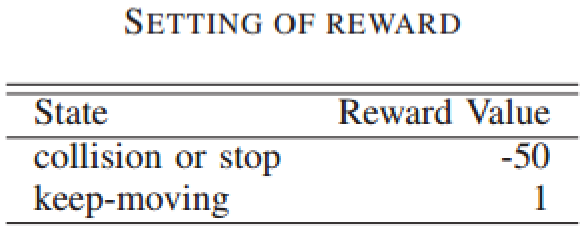
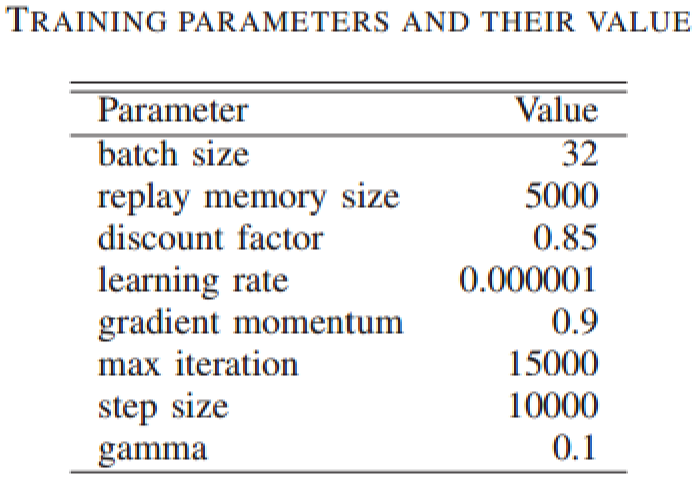


Fig. 2. Feature map extracted from the supervised learning model is the input and is reshaped to a one dimension vector. After three fully-connected hidden layers of a neural network, it will be transformed to the three commands for moving direction as the outputs



SETTING OF REWARD

State	Reward Value
collision or stop	-50
keep-moving	1



TRAINING PARAMETERS AND THEIR VALUE

Parameter	Value
batch size	32
replay memory size	5000
discount factor	0.85
learning rate	0.000001
gradient momentum	0.9
max iteration	15000
step size	10000
gamma	0.1

Lei T, Ming L. A robot exploration strategy based on q-learning network[C]//Real-time Computing and Robotics (RCAR), IEEE International Conference on. IEEE, 2016: 57-62.

Virtual-to-real deep reinforcement learning

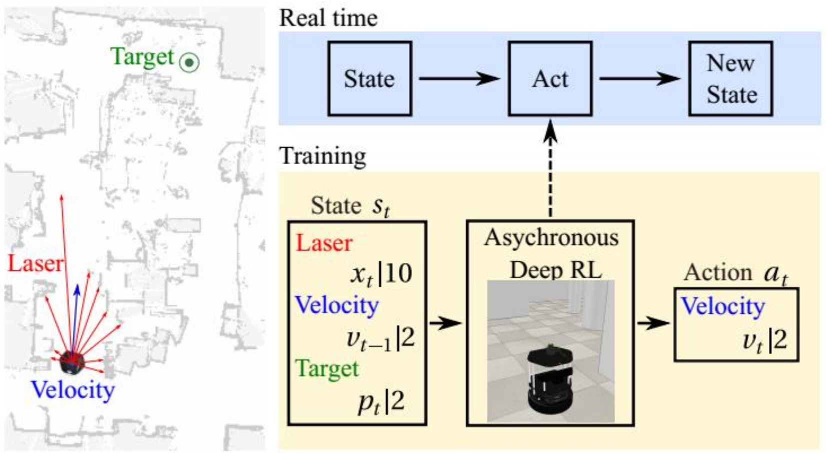


Fig. 1. A mapless motion planner was trained through asynchronous deep-RL to navigate a nonholonomic mobile robot to the target position collision free. The planner was trained in the virtual environment based on sparse 10-dimensional range findings, 2-dimensional previous velocity, and 2-dimensional relative target position.

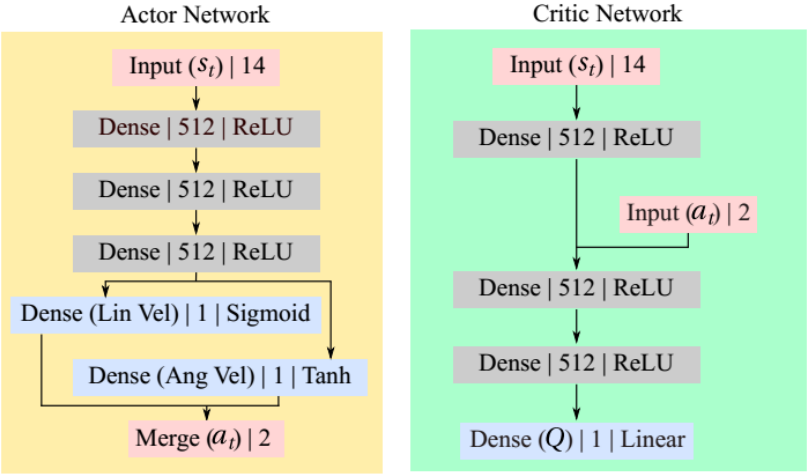


Fig. 3. The network structure for the ADDPG model. Every layer is represented by its type, dimension and activation mode. Notice that the *Dense* layer here means a fully-connected neural network. The *Merge* layer simply combines the several input blobs into a single one.

Tai L, Paolo G, Liu M. Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation[C]//Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on. IEEE, 2017: 31-36.

实现细节

奖励值函数：

$$r(s_t, a_t) = \begin{cases} r_{arrive} \text{ if } d_t < c_d \\ r_{collision} \text{ if } min_{x_t} < c_o \\ c_r(d_{t-1} - d_t) \end{cases}$$

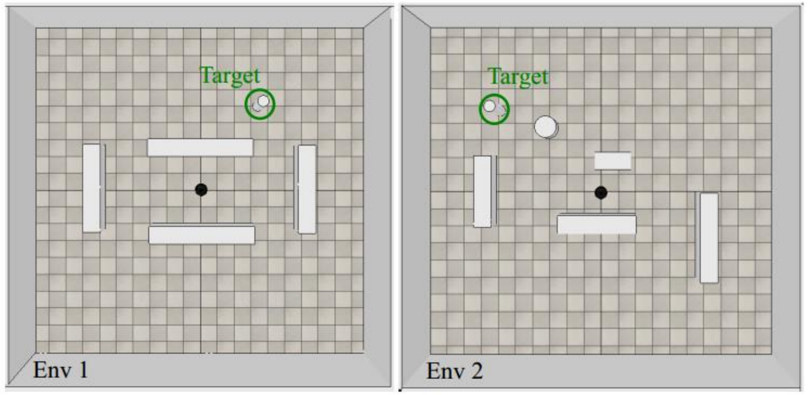
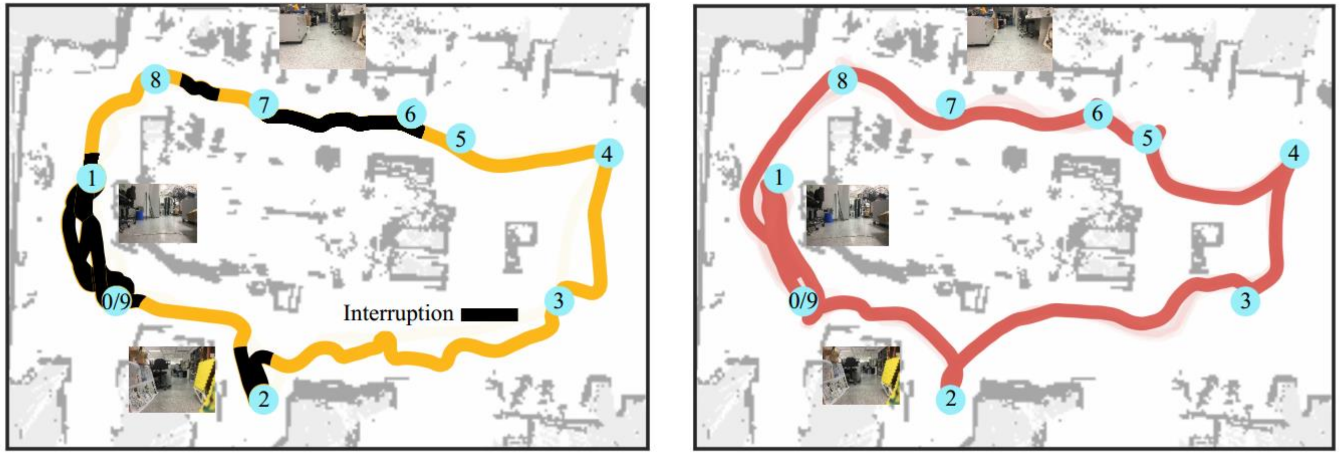


Fig. 4. The virtual training environments were simulated by *V-REP* [21]. We built two 10 × 10 m^2 indoor environments with walls around them. Several different shaped obstacles were located in the environments. A *Turtlebot* was used as the robotics platform. The target labeled in the image is represented by a cylinder object for visual purposes, but it cannot be rendered by the laser sensor. *Env-2* is more compact than *Env-1*.



(a) 10-dim *Move Base*

(b) *Env-2*

Fig. 9. Trajectory tracking in the real test environment. 10-dimensional *Move Base*, and the deep-RL trained model in *Env-2* are compared. 10-dimensional *Move Base* was not able to finish the navigation tasks. Human innervations were added labled as black segments in Fig. 9(a).

Tai L, Paolo G, Liu M. Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation[C]//Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on. IEEE, 2017: 31-36.

A Deep Q Network for Robotic Planning from Image

State	48*48*3，距离地面4m架设的RGB相机拍摄得到的图像
Action	离散化的三个动作（左转、直行、右转）
Env	Gazebo中搭建的室内环境
Reward	-5(collision)-1(move left/right)+10(arrive goal)-0.05(per step)

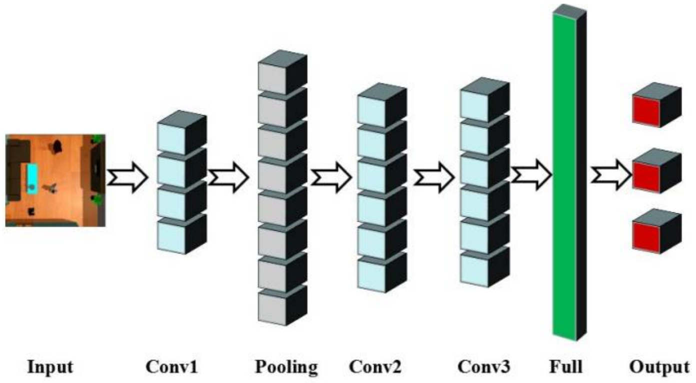


Figure 5: Structure of deep Q network.

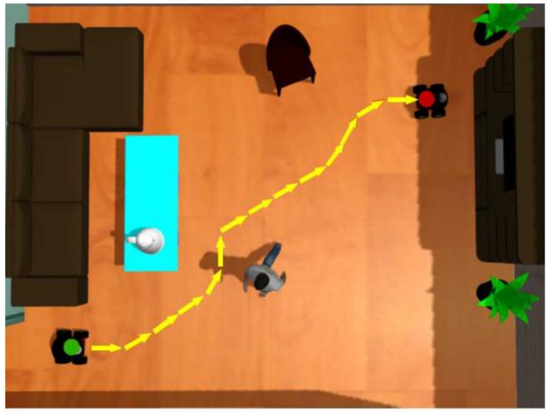
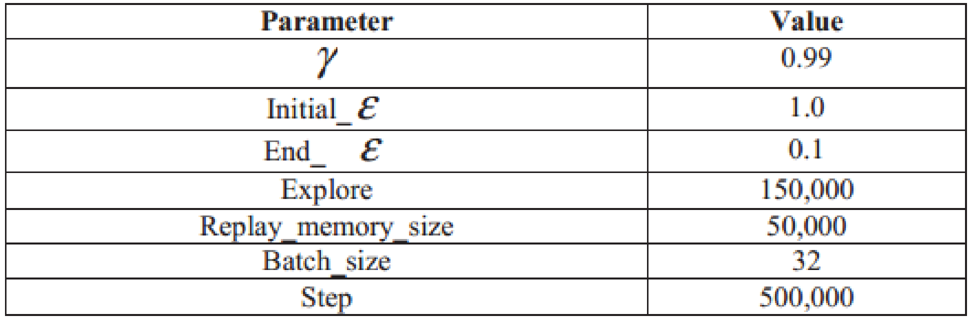


Figure 10: The trajectory of robot movement in the living room.

Han J, Liu H, Wang B. A deep Q network for robotic planning from image[C]//Advanced Robotics and Mechatronics (ICARM), 2017 2nd International Conference on. IEEE, 2017: 626-631.

Application of Deep Reinforcement Learning in Mobile Robot Path Planning

State	40*40，从环境中得到的RGB图像
Action	离散化的三个动作（左转、直行、右转）
Env	DeepMind Lab中的seekavoid_arena_01)
Reward	+1(reach apple)-1(reach lemon)



Parameter	Value
γ	0.99
Initial_ ε	1.0
End_ ε	0.1
Explore	150,000
Replay_memory_size	50,000
Batch_size	32
Step	500,000

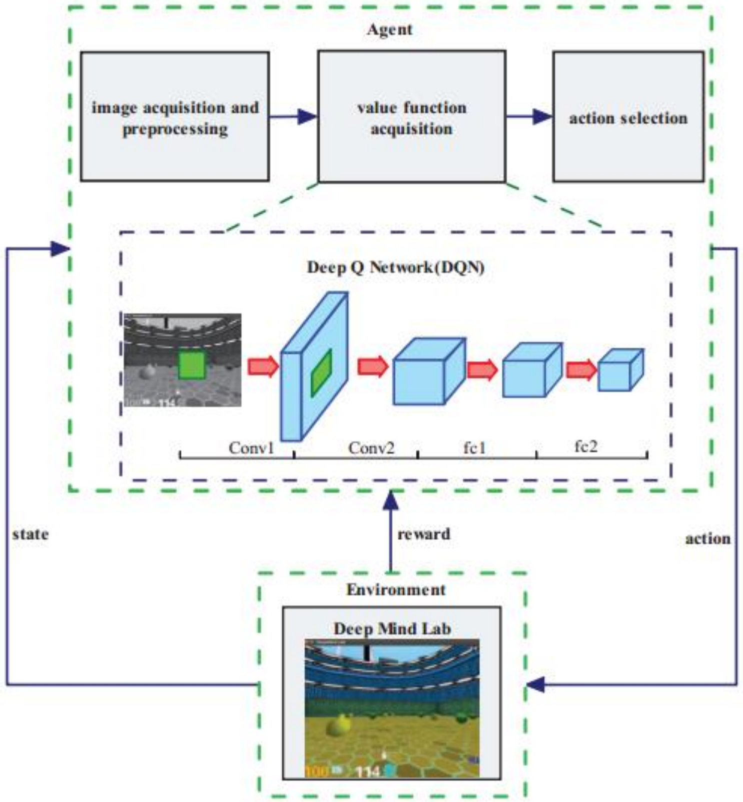
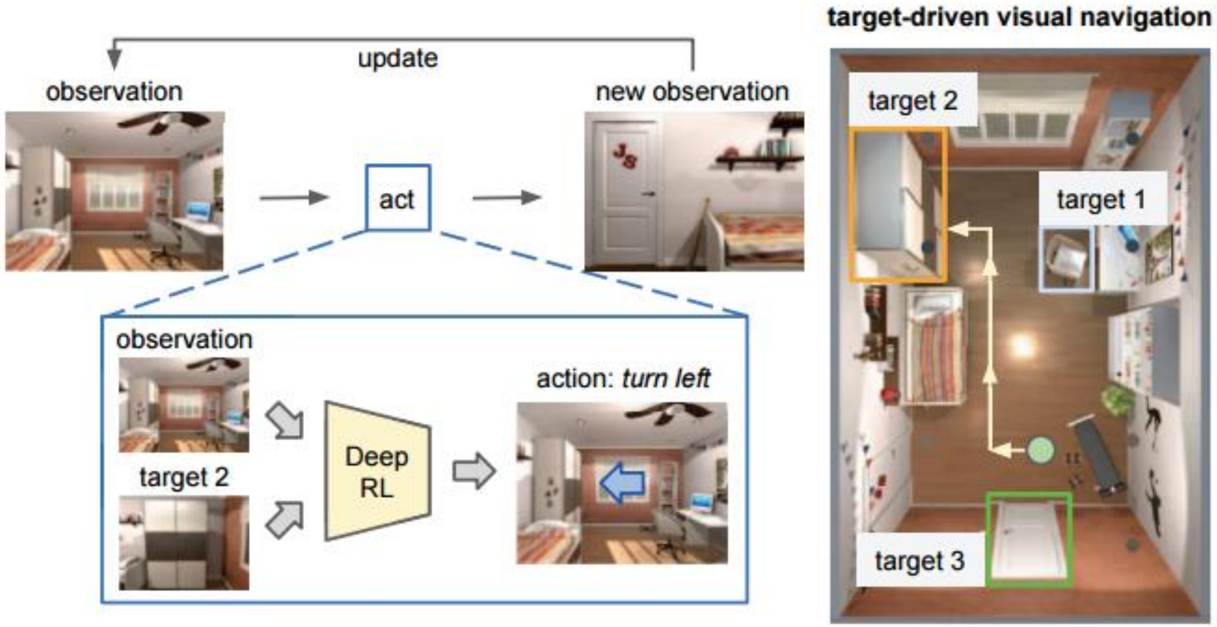


Fig. 1. General framework of mobile robot path planning using deep reinforcement learning

Xin J, Zhao H, Liu D, et al. Application of deep reinforcement learning in mobile robot path planning[C]//Chinese Automation Congress (CAC), 2017. IEEE, 2017: 7112-7116.

Target-driven Visual Navigation in Indoor



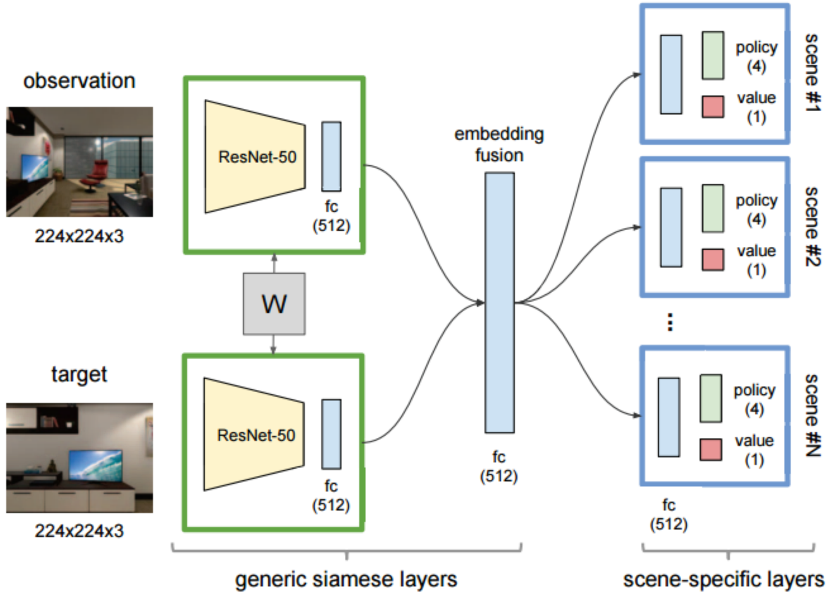
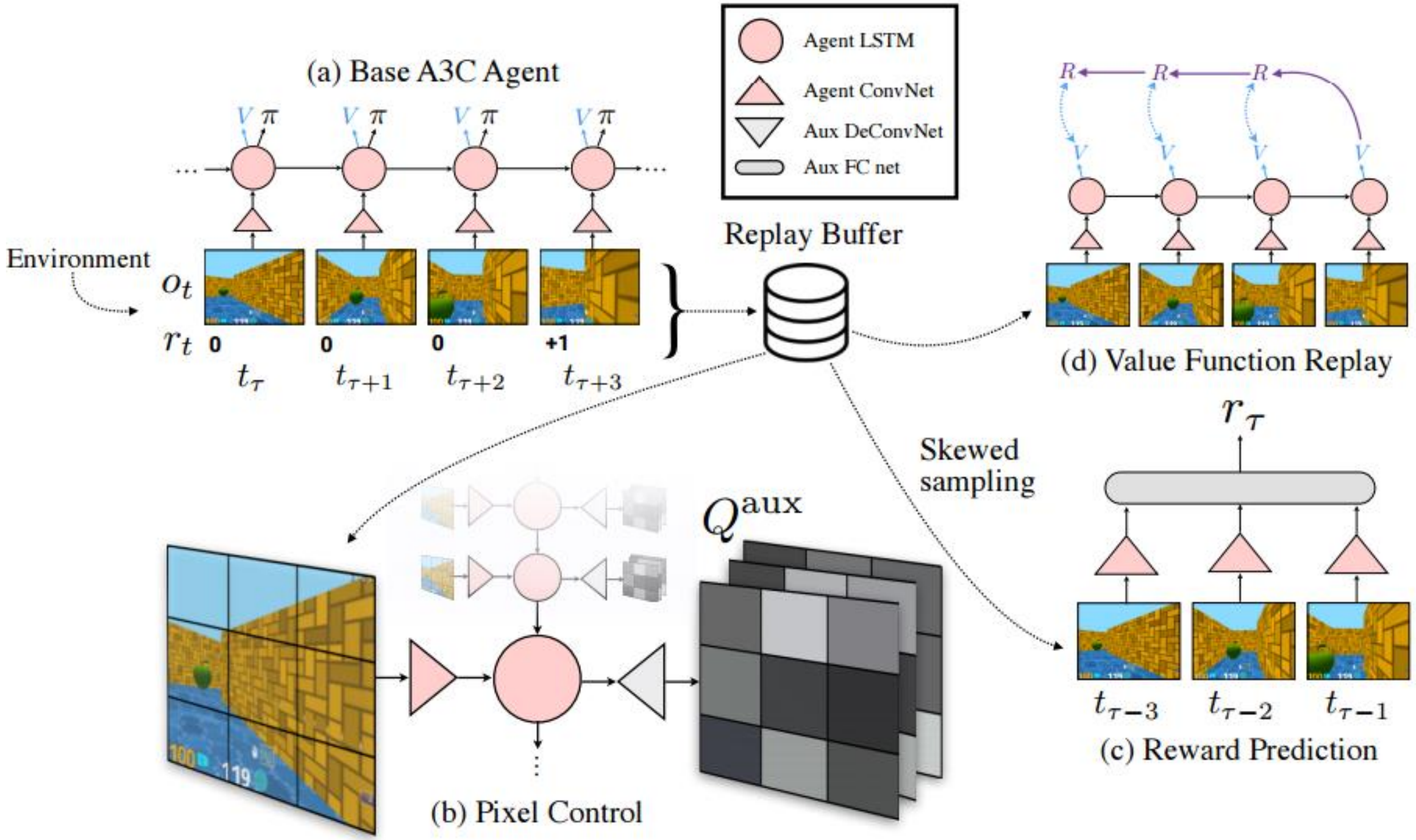


Fig. 4. Network architecture of our deep siamese actor-critic model. The numbers in parentheses show the output dimensions. Layer parameters in the green squares are shared. The ResNet-50 layers (yellow) are pre-trained on ImageNet and fixed during training.

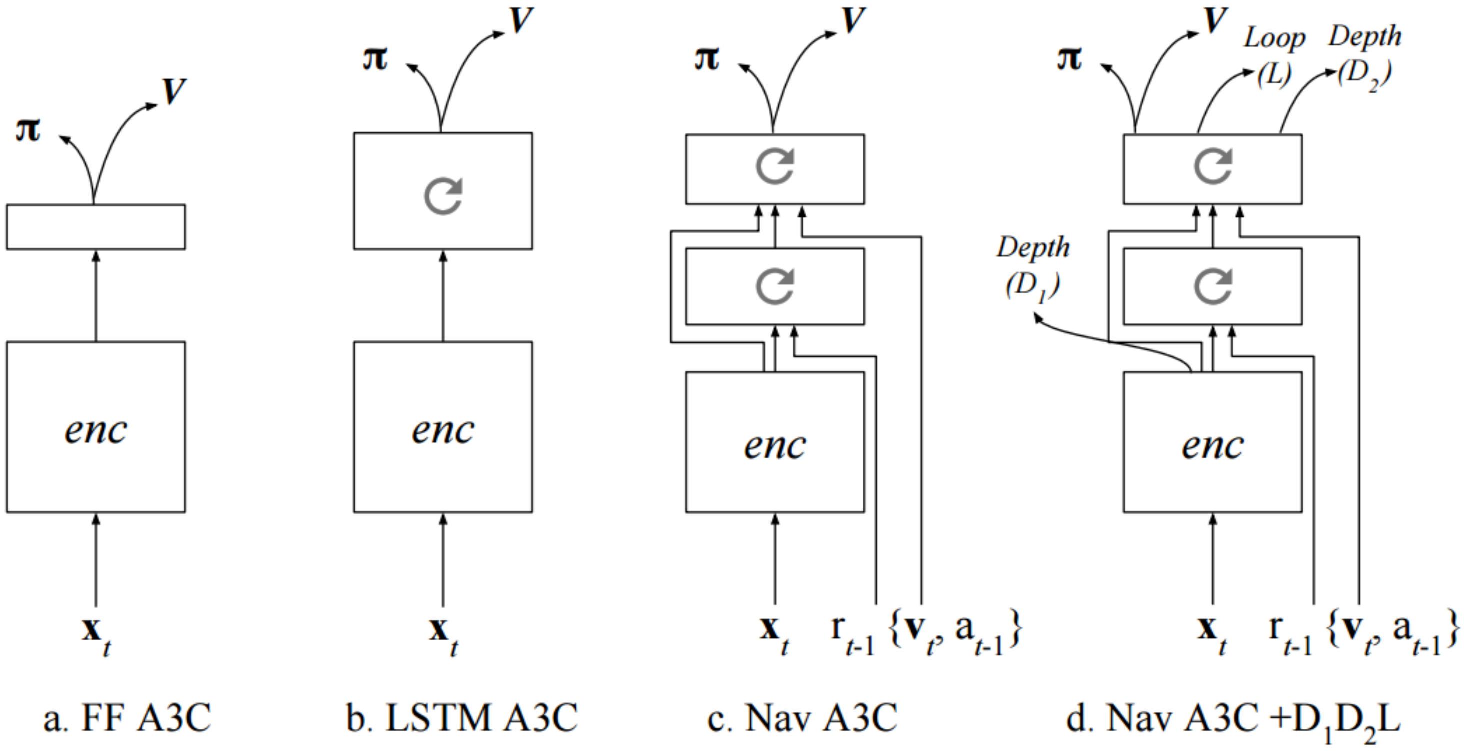
Zhu Y, Mottaghi R, Kolve E, et al. Target-driven visual navigation in indoor scenes using deep reinforcement learning[C]//Robotics and Automation (ICRA), 2017 IEEE International Conference on. IEEE, 2017: 3357-3364.

reinforcement learning with unsupervised auxiliary tasks



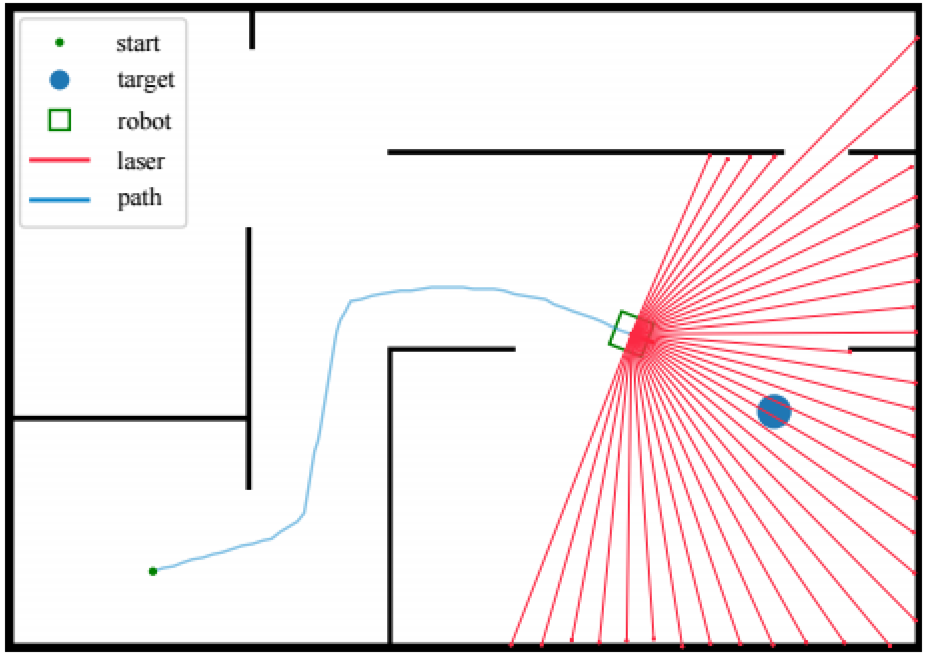
Xin J, Zhao H, Liu D, et al. Application of deep reinforcement learning in mobile robot path planning[C]//Chinese Automation Congress (CAC), 2017. IEEE, 2017: 7112-7116.

Learning to navigate in complex environments



Mirowski P, Pascanu R, Viola F, et al. Learning to navigate in complex environments[J]. arXiv preprint arXiv:1611.03673, 2016.

Curiosity-driven Exploration for Mapless Navigation



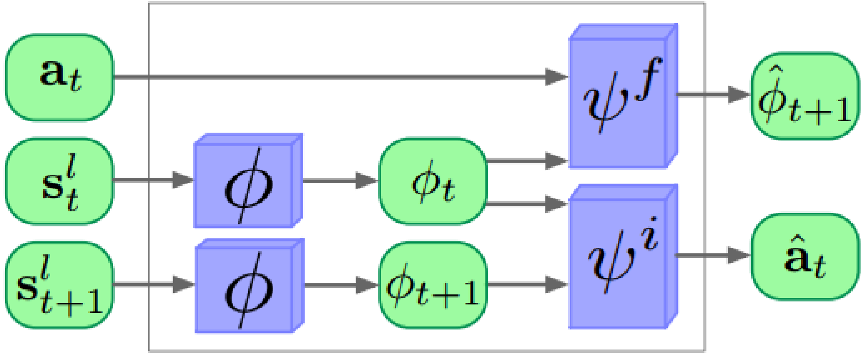
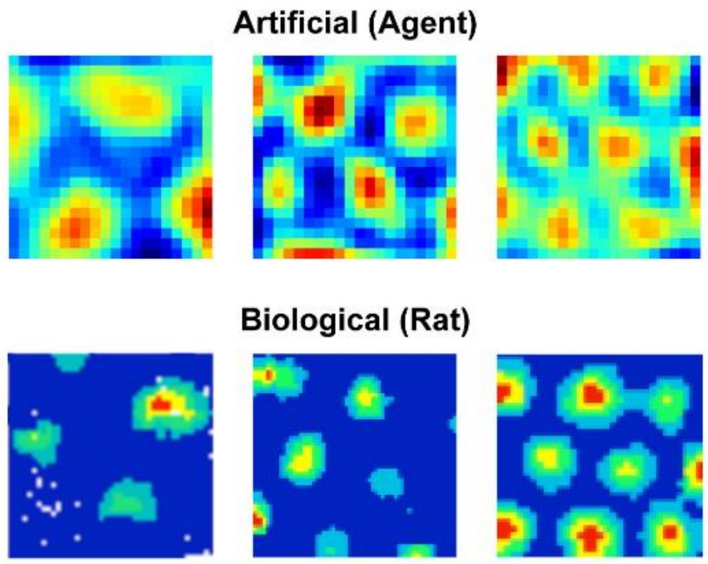


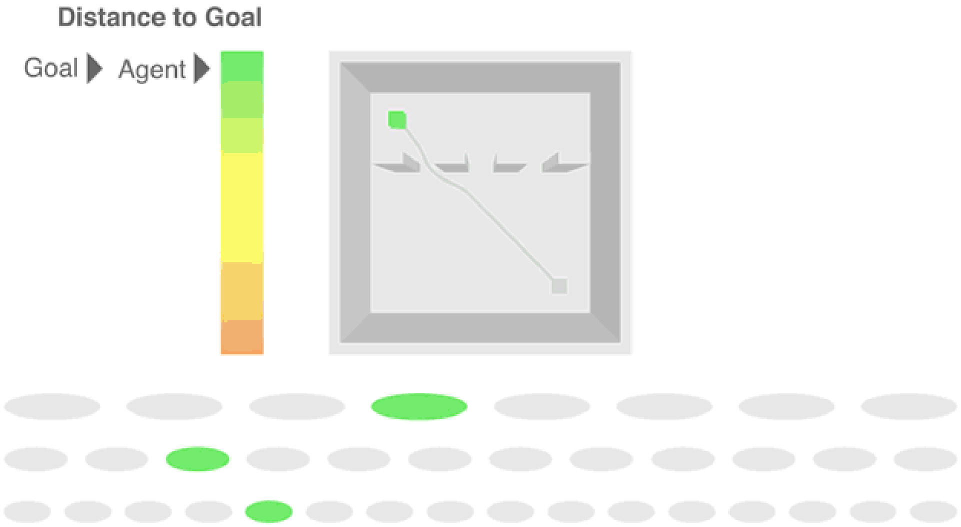
Fig. 2: ICM architecture. $\mathbf{s}_t^l$ and $\mathbf{s}_{t+1}^l$ are first passed through the feature extraction layers ϕ, and encoded into ϕ_t and ϕ_{t+1}. Then ϕ_t and ϕ_{t+1} are input together into the *inverse model* ψ^i, to infer the action $\hat{\mathbf{a}}_t$. At the same time, $\mathbf{a}_t$ and ϕ_t are together used to predict $\hat{\phi}_{t+1}$, through the *forward model* ψ^f. The prediction error between $\hat{\phi}_{t+1}$ and ϕ_{t+1} is used as the intrinsic reward R^i.

Zhelo O, Zhang J, Tai L, et al. Curiosity-driven Exploration for Mapless Navigation with Deep Reinforcement Learning[J]. arXiv preprint arXiv:1804.00456, 2018.

Vector-based navigation using grid-like representations in artificial agents



Our experiments with artificial agents yielded grid-like representations ("grid units") that were strikingly similar to biological grid cells in foraging mammals.



Agent has reached the goal using **vector-based navigation.**

Banino A, Barry C, Uria B, et al. Vector-based navigation using grid-like representations in artificial agents[J]. Nature, 2018: 1.

Learning to Navigate in Cities Without a Map

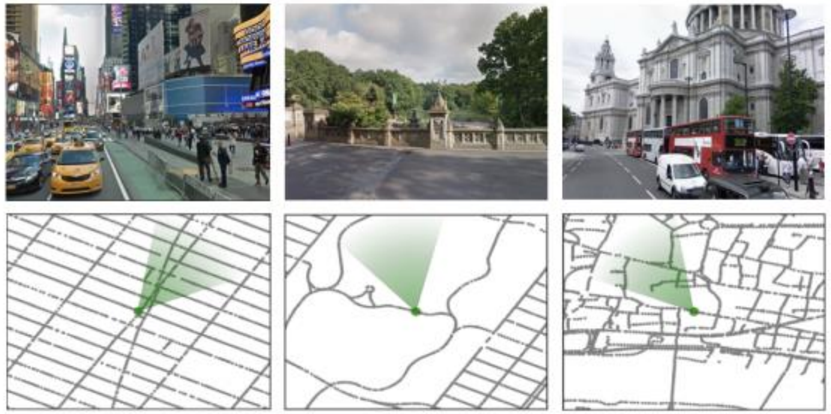
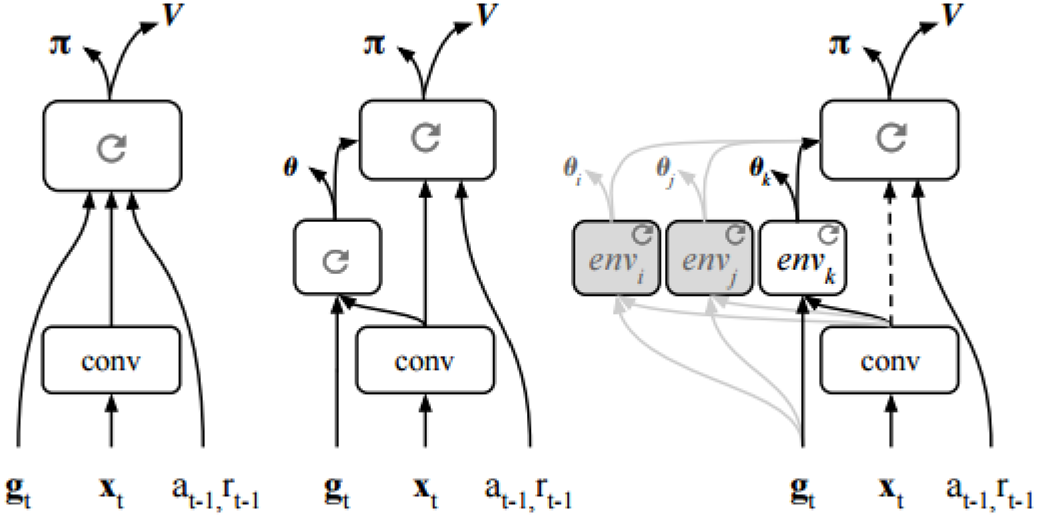


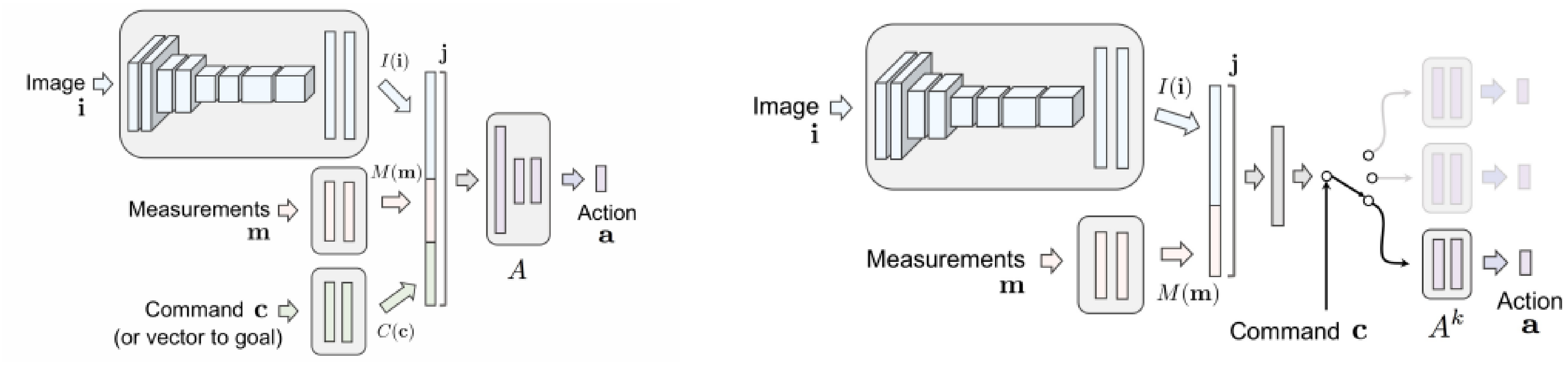
Figure 1. Our environment is built of real-world places from StreetView. The figure shows diverse views and corresponding local maps in New York City (Times Square, Central Park) and London (St. Paul's Cathedral). The green cone represents the agent's location and orientation.



a. *GoalNav* agent b. *CityNav* agent c. *MultiCityNav* agent

Mirowski P, Grimes M K, Malinowski M, et al. Learning to Navigate in Cities Without a Map[J]. arXiv preprint arXiv:1804.00168, 2018.

End-to-end driving via conditional imitation learning



(a)

(b)

Fig. 3. Two network architectures for command-conditional imitation learning. (a) `command input`: the command is processed as input by the network, together with the image and the measurements. The same architecture can be used for goal-conditional learning (one of the baselines in our experiments), by replacing the command by a vector pointing to the goal. (b) `branched`: the command acts as a switch that selects between specialized sub-modules.

Codevilla F, Müller M, Dosovitskiy A, et al. End-to-end driving via conditional imitation learning[J]. arXiv preprint arXiv:1710.02410, 2017.