گرفتن اجسام دارای مفصل کشویی و یا لولایی با استفاده از ربات دو انگشتی و به کمک شبکه های عمیق و یادگیری تعاملی

دانشجو: **احسان فروتن** ehsan.forootan@ut.ac.ir

استاد راهنما: **دکتر مهدی طالع ماسوله** m.t.masouleh@ut.ac.ir

گروه کنترل دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران (ECE Department, University of Tehran)

مقدمه

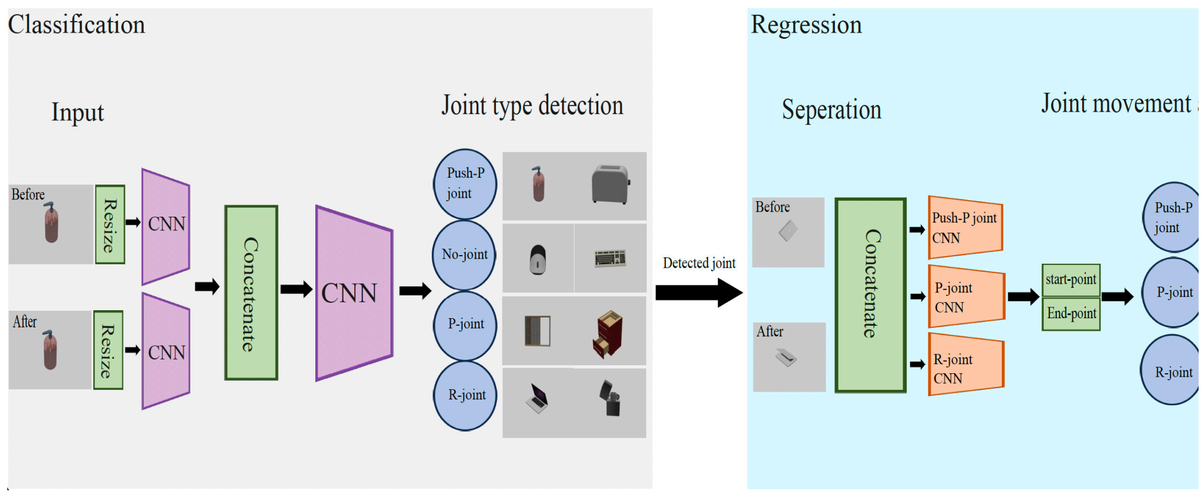
به صورت خلاصه در این پروژه، دو موضوع بررسی خواهد شد.

- **اولین موضوع**، بررسی تشخیص نوع و نحوه حرکت اجسام دارای مفصل قبل و بعد از تعامل با آنها، بر روی مجموعه داده **PartNet-Mobility** با بهره گیری از **شبکه های کانولوشنی عمیق** است. در اینجا، ساختاری ارائه شده است که در **تشخیص نوع جسم** به **دقت ۹۶ درصد** و در **تشخیص نحوه حرکت** به **دقت ۹۴ درصد** رسیده است.

- **موضوع دوم** که در این پروژه مورد بررسی است، تعامل و باز کردن اجسام دارای مفصل در **محیط شبیه سازی شده ManiSkill2** است. در اینجا، با کمک بهره گیری از **معماری CNN+LSTM+Attention** و یادگیری با **روش Actor-Critic**، روشی ارائه شده است که در باز کردن درب اجسام دارای **مفصل کشویی** به **دقت ۵۵.۰ درصد** و در باز کردن درب اجسام دارای **مفصل لولایی** به **دقت ۴۹.۷ درصد** می رسد.

روش تشخیص نوع و حرکت اجسام مفصلی

در این روش، ساختار زیر، با بهره گیری از ساختار موجود در شبکه VGG، پیشنهاد شده است:

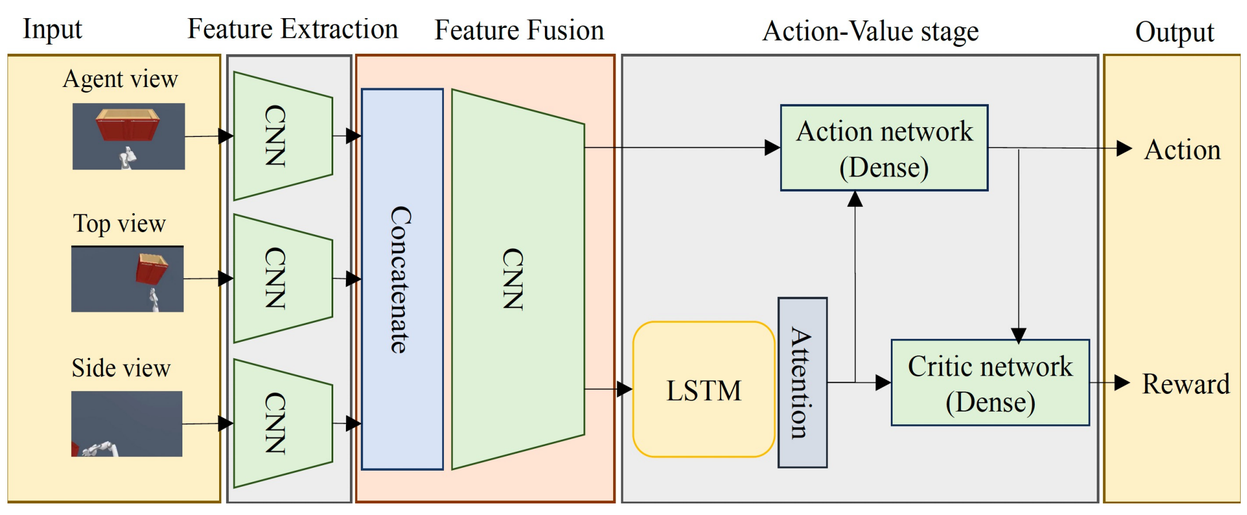


- در این ساختار، بخش اول برای **تشخیص نوع مفصل** با استفاده از **دو تصویر قبل و بعد از تعامل با جسم مفصلی** به کار می رود. **هر تصویر** از **یک شبکه** عبور داده می شود تا ماتریس های ویژگی متناظر استخراج شود. سپس این ویژگی ها به **یک شبکه سوم** داده می شود تا **طبقه بندی** انجام گردد.

- پس از تشخیص نوع مفصل، یک شبکه برای رگرسیون نقطه شروع و پایان استفاده می شود. در این بخش از معماری، به ازای هرنوع جسم با مفصل لولایی، کشویی ویا کشویی فشاری یک شبکه با ساختار یکسان اما وزن های به خصوص خود استفاده شده است تا تشخیص نحوه حرکت با دقت هرچه بیشتر انجام بپذیرد.

روش تعامل با اجسام مفصل دار

برای تعامل با اجسام دارای مفصل، ساختار زیر در این پروژه پیشنهاد شده است:



- در ساختار پیشنهاد شده، **سه نما** از محیط به **سه لایه کانولوشنی مجزا** داده می شوند. **خروجی** این لایه ها با هم **ترکیب شده** و از یک **لایه کانولوشنی دیگر** عبور کرده و سپس وارد **یک LSTM** و پس از آن، وارد **یک لایه Attention** می شوند.

- در نهایت، حالت ورودی پس از پردازش از لایه های گفته شده، وارد **شبکه های Actor و Critic** می شوند تا **پیش بینی Action و Reward** را انجام دهند و ربات حرکت کند.

نتایج تشخیص نوع و حرکت اجسام مفصلی

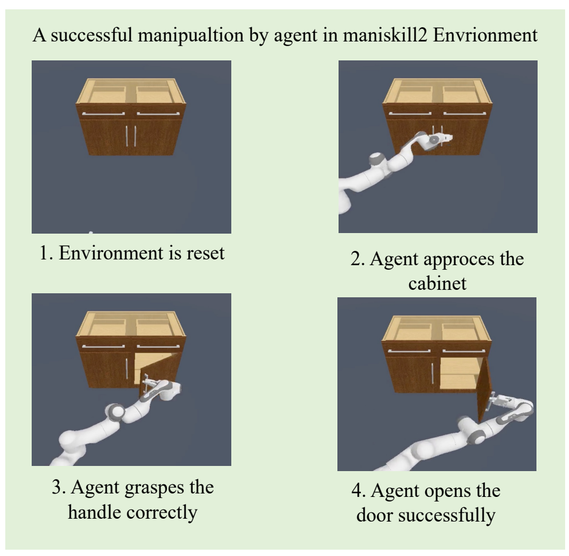
جدول زیر، دقت روش پیشنهادی برای تشخیص نوع و حرکت اجسام مفصلی در **PartNet-Mobility dataset** را در مقایسه با روش های موجود در ادبیات را به نمایش می گذارد.

Method	Regression (MAEP)	Classification (F1-Score)	Total MSE
SVM	77.66%	92%	0.345
MLP	70.66%	82%	0.505
ResNet50-V2	85.67%	93%	0.156
JoTyMo (Joint Type And Movement Detection)	**94%**	**96%**	**0.085**

نتایج تعامل با اجسام مفصل دار

جدول زیر، درصد موفقیت در تعامل با اجسام مفصلی در محیط Maniskill2 را برای معماری پیشنهادی در مقایسه با روش های موجود در ادبیات را به نمایش می گذارد.

Method	Door	Drawer	Total
Maniskill2 Baseline	30.0%	37.0%	33.5%
DDPG	32.0%	30.0%	31.0%
PPO	33.0%	35.0%	34.0%
DQN	24.0%	28.0%	26.0%
PartManip	49.1%	54.7%	51.9%
CNN+LSTM+Attention	**49.7%**	**55.0%**	**52.25%**



نمونه نحوه گرفتن یک جسم دارای مفصل در این پروژه

جمع بندی

- **متد اول** مطرح شده در این پروژه با بهره گیری از **شبکه های کانولوشنی**، به **دقت ۹۶ درصد** در **تشخیص نوع جسم مفصلی** و به **دقت ۹۴ درصد** در **تشخیص نحوه حرکت جسم مفصلی** دست یافت.
- **متد دوم** مطرح شده در این پروژه با استفاده از **معماری مبتنی بر CNN+LSTM+Attention** توانست در **۴۹.۷ درصد** مواقع با **موفقیت** با اجسام **لولایی** و در **۵۵.۰ درصد** مواقع با **موفقیت** با اجسام **کشویی** تعامل و آنها را باز بکند. بنابراین **درصد موفقیت کلی** در این متد، **۵۲.۲۵ درصد** است که در مقایسه با روش های موجود در ادبیات به موفقیت بیشتری دست می یابد.

مراجع اصلی

[1] E. Forootan, H. Ghasemi, M. T. Masouleh, A. Kalhor and B. Moshiri, "JoTyMo: Joint Type and Movement Detection from RGB Images of Before and After Interacting with Articulated Objects," *2023 11th RSI International Conference on Robotics and Mechatronics (ICRoM)*, Tehran, Iran, Islamic Republic of, 2023, pp. 592-597, doi: 10.1109/ICRoM60803.2023.10412492. keywords: {Convolutional codes;Training;Transfer learning;Network architecture;Convolutional neural networks;Task analysis;Testing},

[2] E.Forootan ,A.Rashidi ,H.Ghasemi ,M.T.Masouleh ,A.Kalhor and B.Moshiri,"LSTM-Attention based Actor-Critic reinforcement learning for articualted object manipulation ," unpublished

Human and Robot Interaction Laboratory
Department of Electrical and Computer Engineering
University of Tehran
Tehran, Iran