CS686 MPA 2nd Paper Presentation

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement

(RAL 2022)

2023.11.27.

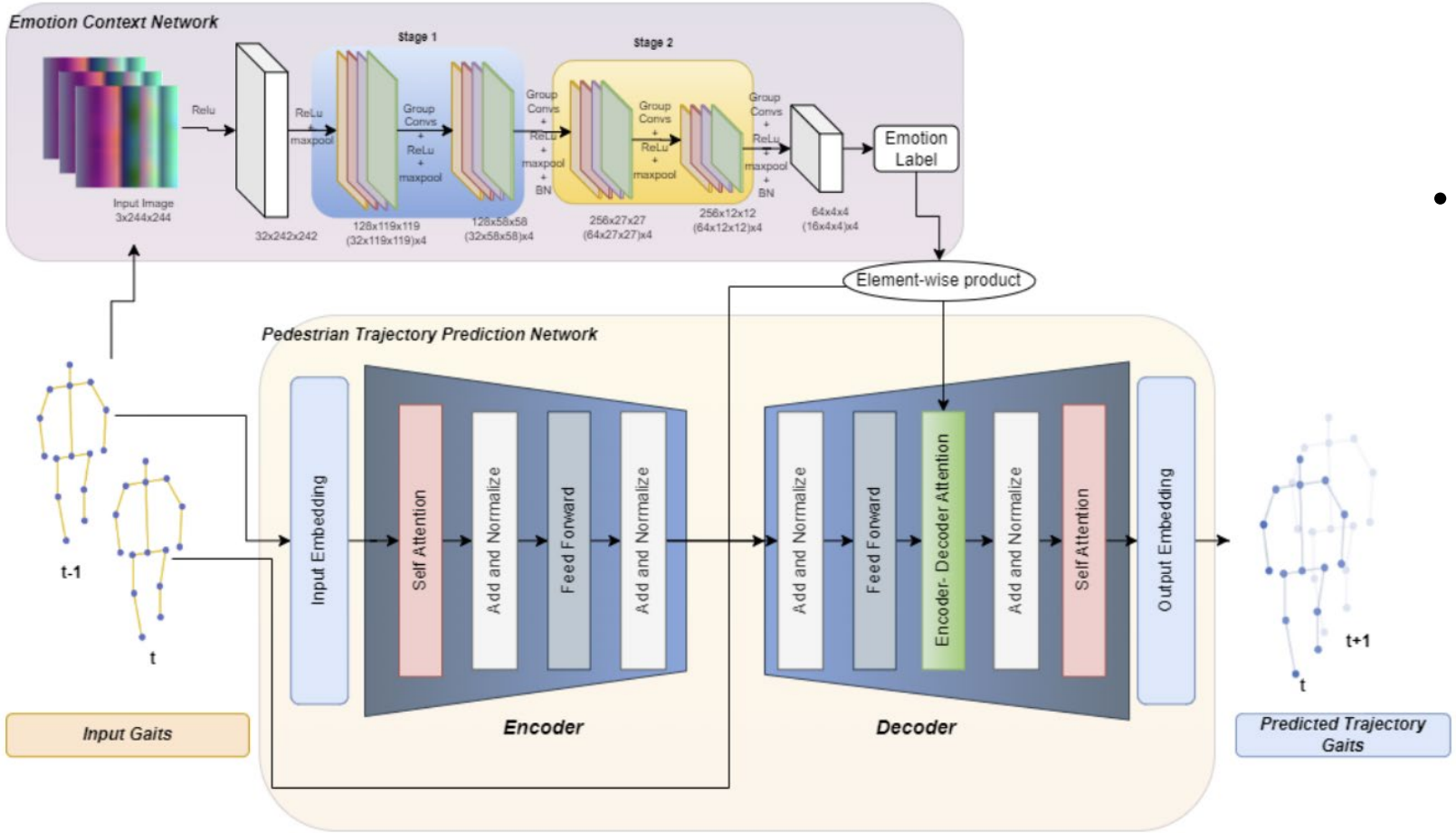
1st year M.S. Student @KAST, SGVR Lab
CHANMI LEE

KAIST

Review

EWareNet: Emotion-Aware Pedestrian Intent Prediction and Adaptive Spatial Profile Fusion for Social Robot Navigation

(explained by Jihyeok)



- Pedestrian Intent Prediction
  - Pose Extraction
  - Emotional Context Net
  - Pedestrian Trajectory Prediction Net

➔ Predict Trajectory Gaits

➔ Intent-Aware Navigation

KAIST

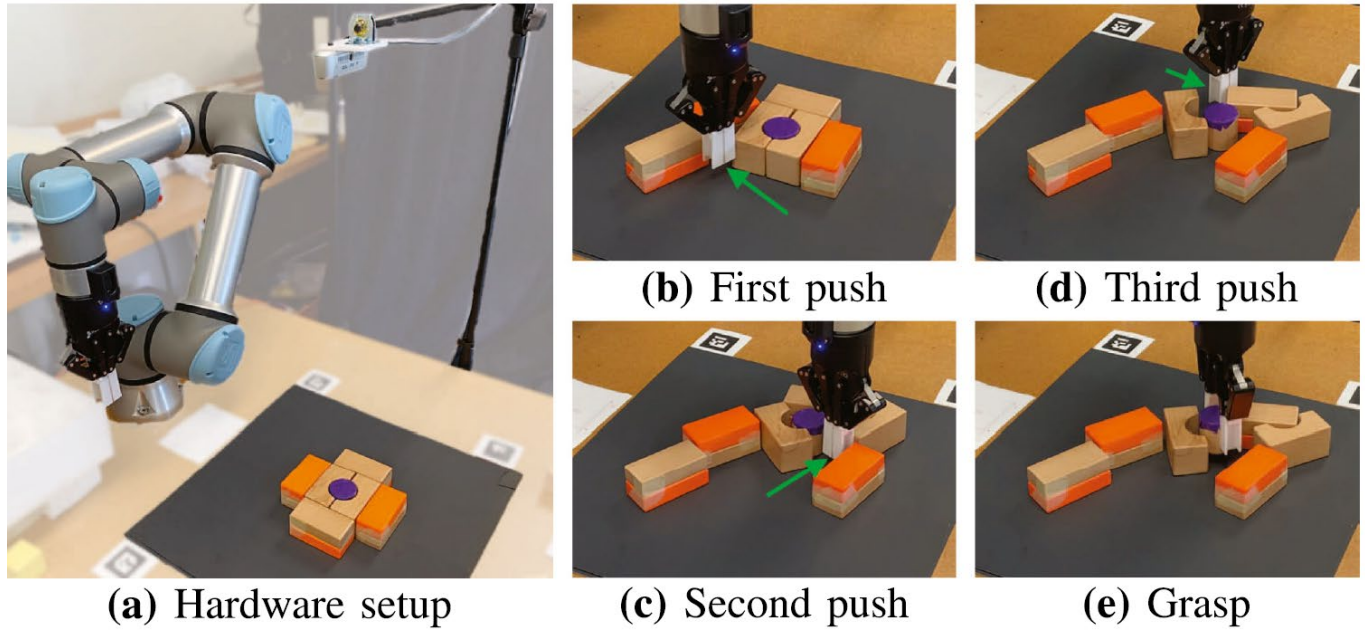
Contents

Overview

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement

Purpose & Main Idea

- Grasp object retrieval in dense clutter.
  - Uses Visual Foresight Tree (VFT)
  - Deep neural network
    ➔ predict object poses from pushes
  - Tree search algorithm
    ➔ plan optimal sequence of pushes
  - Improve solution quality



(a) Hardware setup (b) First push (c) Second push (d) Third push (e) Grasp

KAIST

Background

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement

DIPN

- Predict object movements ➔ create post-push synthetic images

  (1) Image segmentation

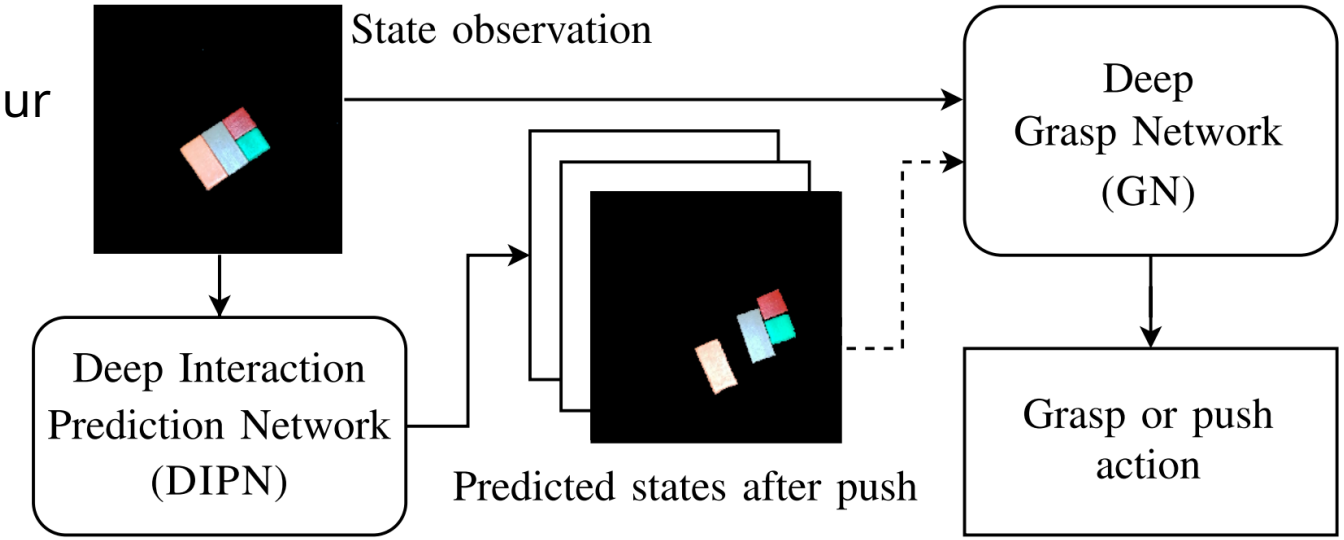
  - Mask R-CNN → mask & center of each object

  (2) Push Sampling

  - sample push action on object contour
  - R-CNN → feasibility check

  (3) Push Prediction

  - Predict object transformations
  - MLP → encode push action
  - ResNet → encode mask image



KAIST

Background

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement

MCTS

- Heuristic algorithm for decision process

  (1) Selection

  - Select best node to grow tree

  (2) Expansion

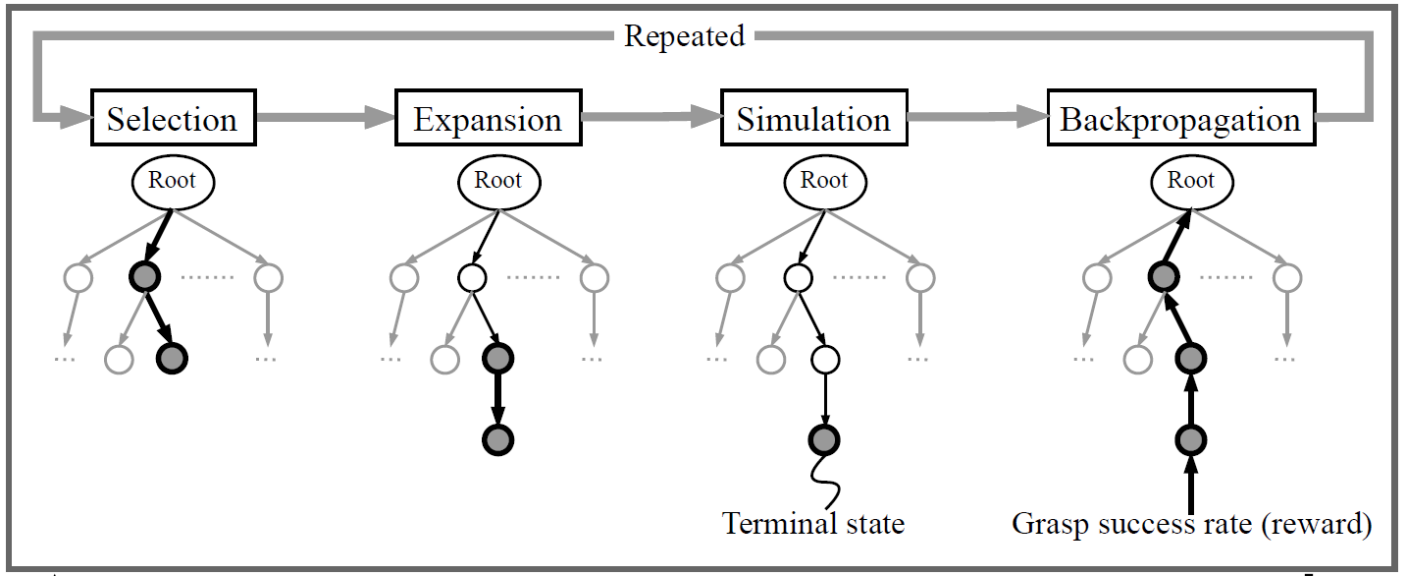
  - Add new child node to the tree

  (3) Simulation

  - Carry out at new child node

  (4) Backpropagation

  - Obtain terminal reward



KAIST

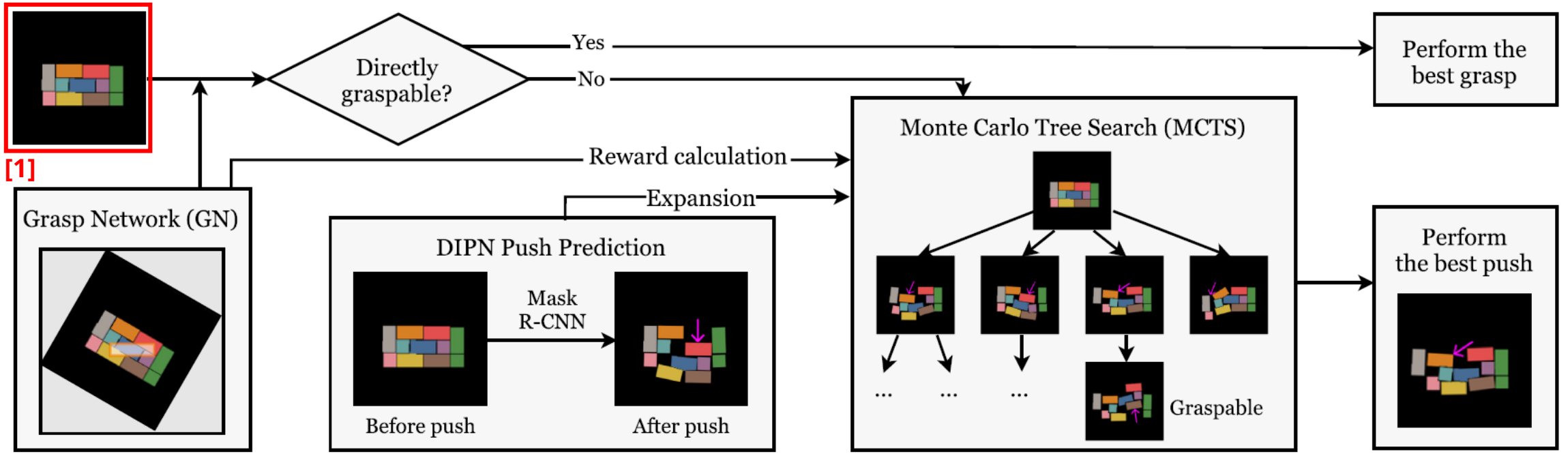
Methodology

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement

[1] Scene Perception

- Take RGB-D image
  - Objects are detected

➔ Classify as unknown clutter or target object



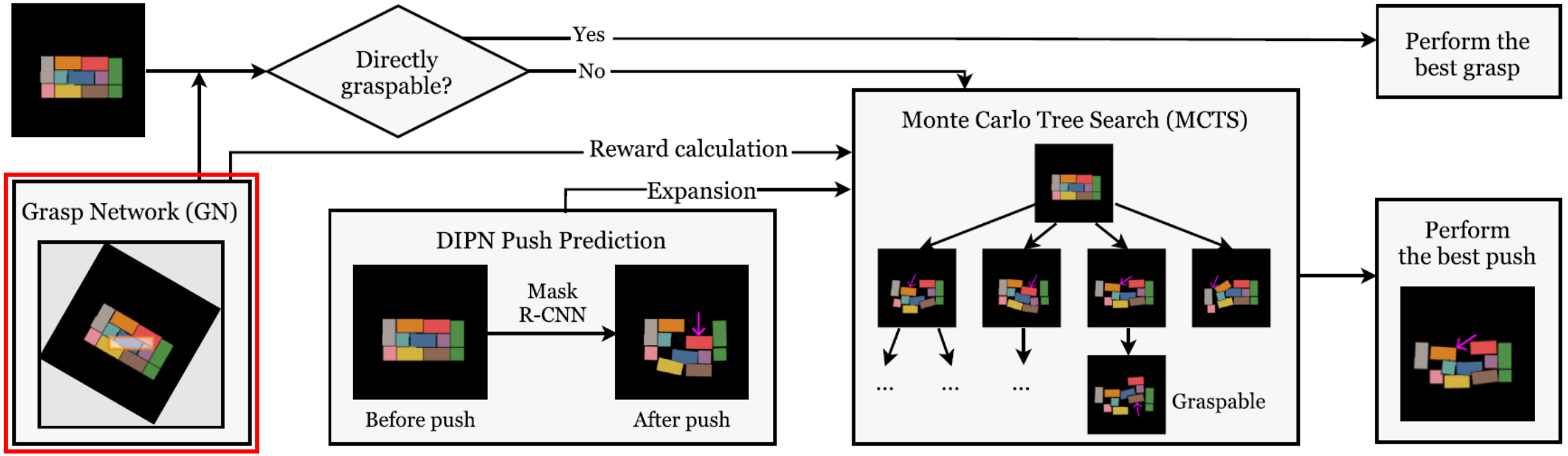
Methodology

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement

[2] Grasp Network

- Estimate grasp Q-value on target object
  - Get grasp probability and use it as a threshold value

➔ Determine whether the target object can be grasped



[2]

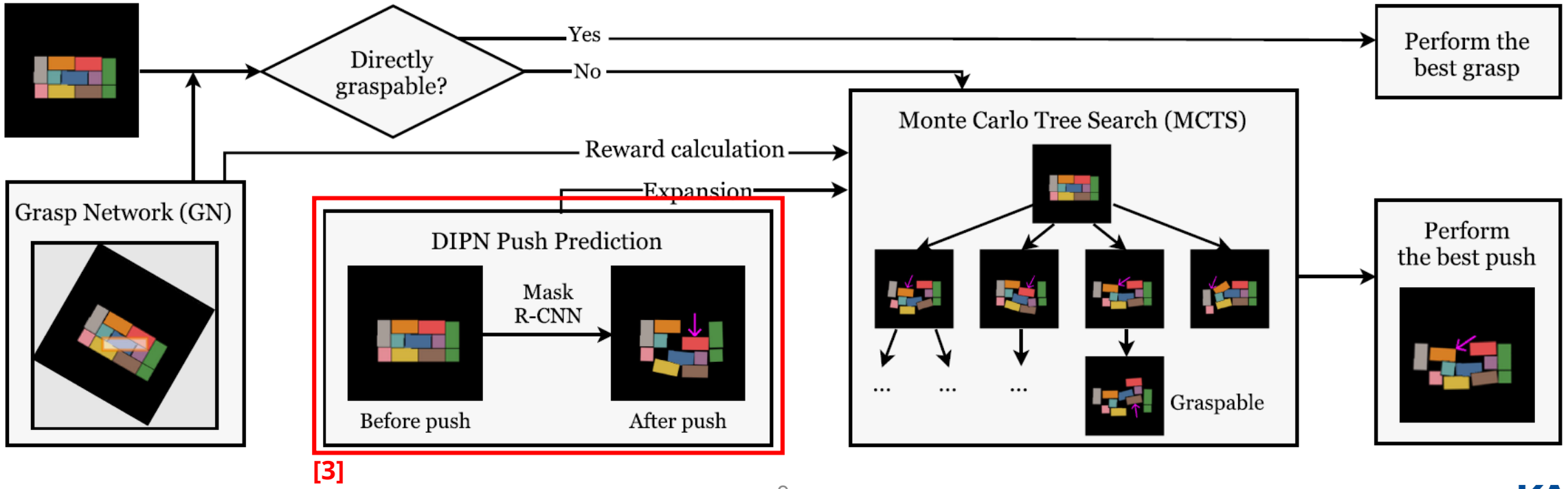
Methodology

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement

[3] Push Prediction using DIPN

- Get post-push synthetic images

  ➔ Generate the predicted next state after a push



[3]

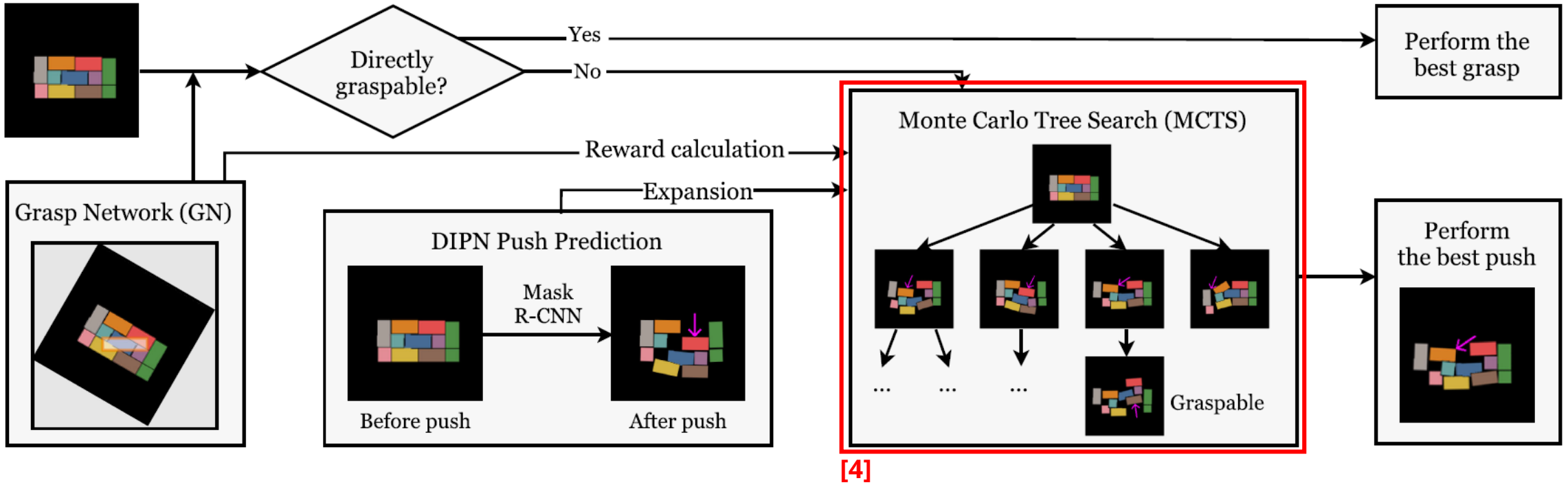
Methodology

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement

[4] Monte Carlo Tree Search (MCTS)

- Iterate “selection-expansion-simulation-backpropagation”
  - Define N_max (maximum number of node) → computational budget to stop the search

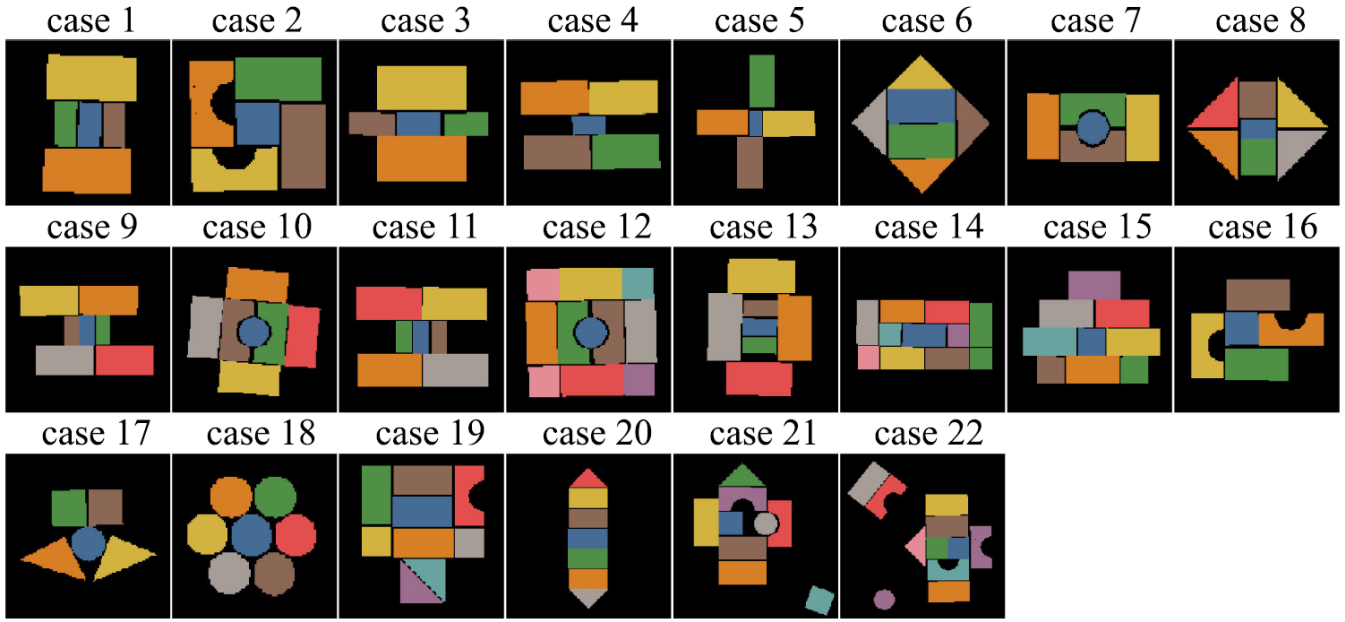
  ➔ Find best push action to take



[4]

Experimental Results

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement



- Test cases used in experiments
  - Target object : Blue block

KAIST

Experimental Results

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement

	Completion	Grasp Success	Number of Actions
gc-VPG [33]	89.3%	41.7%	5.78
go-PGN [33]	99.0%	90.2%	2.77
DIPN [4]	100%	100%	2.30
VFT (ours)	100%	100%	**2.00**

	Completion	Grasp Success	Num. of Actions
DIPN [4]	100%	98.3%	4.31
VFT (ours)	100%	98.8%	**2.45**

	Completion	Grasp Success	Num. of Actions
DIPN [4]	100%	97.0%	4.78
VFT (ours)	100%	98.5%	**2.65**

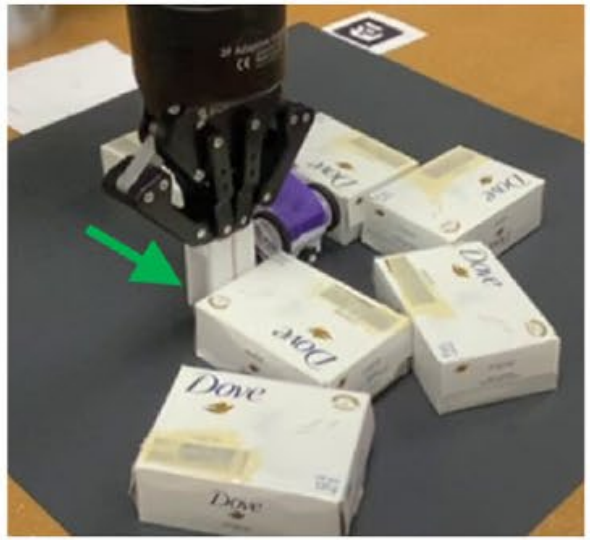
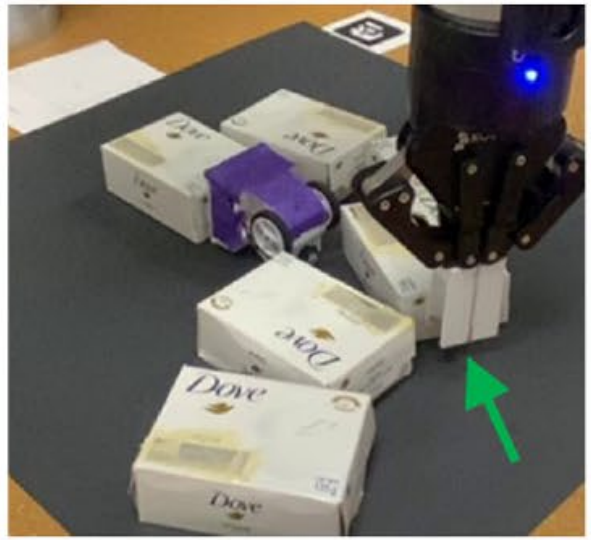
- Simulation results for the 10 test cases
  ➔ VFT is the most efficient
- Simulation result for the 22 test cases
  ➔ VFT is more efficient than DIPN
- Real experiment results for the 22 test cases
  ➔ VFT is more efficient than DIPN

KAIST

Experimental Results

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement

https://www.youtube.com/watch?v=7cL-hmgvyec (5’18” ~)


KAIST

Summary

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement

Conclusion & Discussion

- VFT (Visual Foresight Trees)
  - Scene Perception
  - DIPN
  - MCTS
  - Find best push action
- Strength
  - High quality, multi-horizon prediction
- Weakness
  - Weakness in real world test scenarios

Thank You

Quiz

Visual Foresight Trees for Object Retrieval from Clutter with Nonprehensile Rearrangement

Q1) What is not related with VFT?

A. Image segmentation
B. DIPN
C. MCTS
D. PMBS

Q2) What is the key difference between DIPN and VFT?