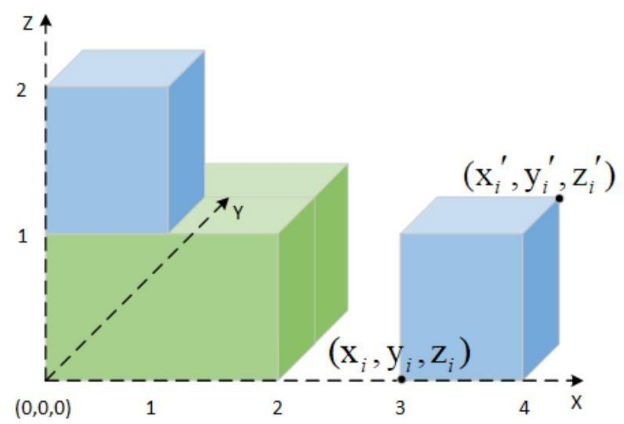
Solving 3D Bin Packing Problem via Multimodal Deep Reinforcement Learning

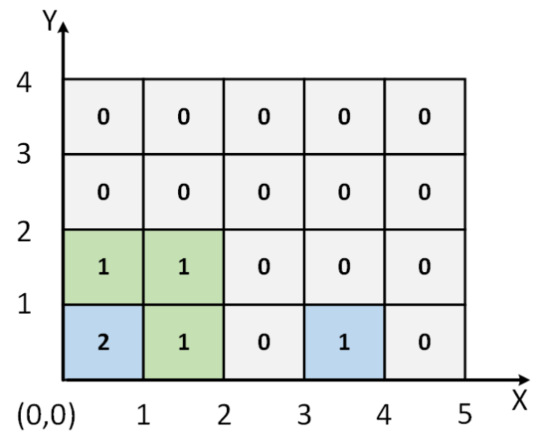
Yuan Jiang[1], Zhiguang Cao[2], Jie Zhang[1]
[1]Nanyang Technological University
[2]National University of Singapore

Motivation

- Recently, there is growing attention on applying deep reinforcement learning (DRL) to solve the 3D bin packing problem (3D BPP).
- However, due to the relatively less informative yet computationally heavy encoder, and considerably large action space inherent to the 3D BPP, existing methods are only able to handle up to 50 boxes.
- We propose to alleviate this issue via an end-to-end multimodal DRL agent, which sequentially addresses three sub-tasks of sequence, orientation and position, respectively.

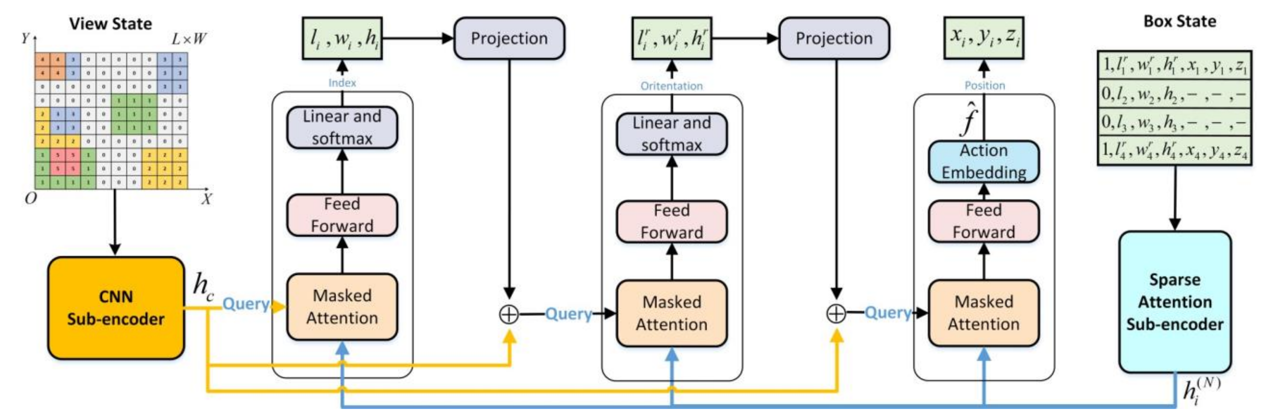
3D Bin Packing Environment



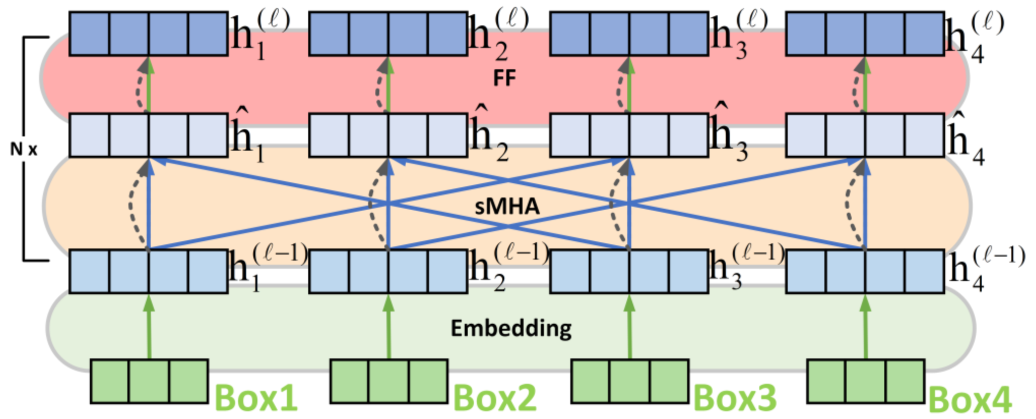


- The objective of 3D BPP is to pack rectangular boxes with 3 dimensions (length, width and height) into the bin, while maintain constraints and minimize the final stacked height.
- We stipulate that the boxes are dropped from the top straightly downwards. The motivation for the view state is to offer more informative spatial features to facilitate the packing.

Method



- Our agent adopts an encoder-decoder diagram to learn the packing policy.
- The multimodal encoder maps the states into feature embeddings, and a decoder is responsible for incrementally constructing solutions for the three sub-tasks.

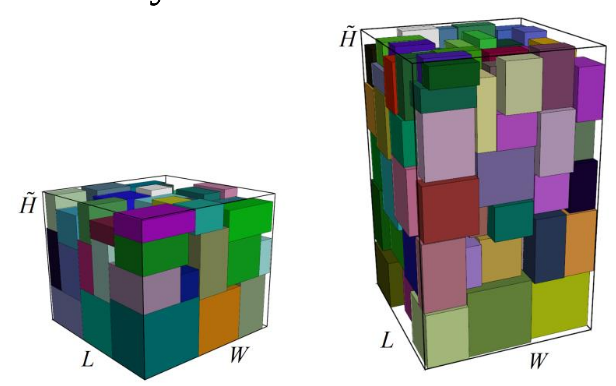


- In the multimodal encoder, a sparse attention sub-encoder is exploited to embed the box state (i.e., the basic information of the box including the indication of whether packed, orientation and position)
- A CNN (convolutional neural network) sub-encoder is used to embed the top-down view for more informative auxiliary representation.
- In the decoder, an action representation learning is leveraged to deal with the large action space that mainly caused by the position sub-task.

RESULTS

Box Number	*GA*	*EP*	*MTSL*	*CQL*	*OUR*
20	68.3%	62.7%	62.4%	67.0%	**71.8%**
30	66.2%	63.8%	60.1%	69.3%	**75.5%**
50	65.9%	66.3%	55.3%	73.6%	**81.3%**
100	62.4%	67.5%	-	-	**84.4%**

- The metric is utilization rate. The superior performance reflected in the table well justified the overall efficacy of the designed in our method for boosting the solution quality and computation efficiency.



- Larger number of boxes often means more information in the input sequence, which could be well exploited by the attention mechanism and DRL algorithm to learn the relationship among boxes and facilitate the long-term planning.

CONCLUTION

- We employ a multimodal encoder with a sparse attention sub-encoder and a CNN sub-encoder to exploit multimodal information. Meanwhile, the action representation learning is adopted to cope with large action space. The resulting policy enables the agent to solve large instances of 100 boxes or more. Moreover, our method also delivers superior performance in terms of utilization rate against all the baselines.