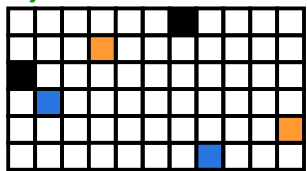


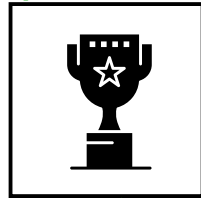
Control Plane

5) Action



4) Flatten Observation

7) Reward



Reward

3) Predicted CPU of Nodes

3) YOLO Profiled Data

3) Cluster State

6) Container Placement and Model Switch Action in GKE/Simulated Testing

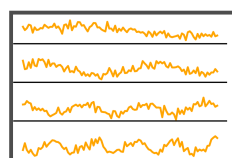
6) Container Placement and Model Switch Action in Training

IoT Layer

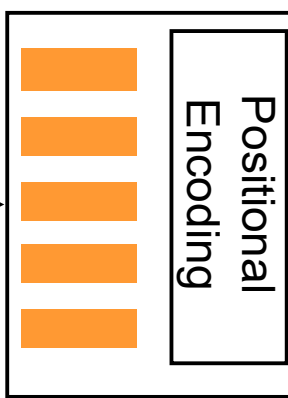
8) Public Traffic

PatchTST

1) Nodes CPU



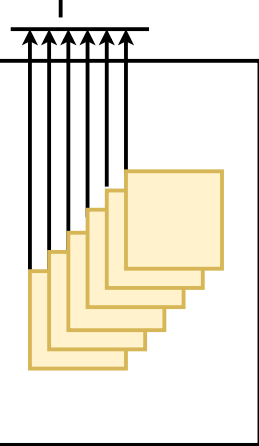
2a) Patching



2a) Vectors



2a) Transformers

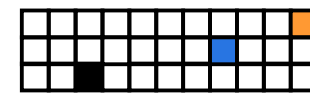
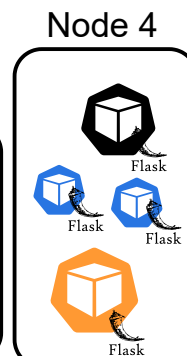
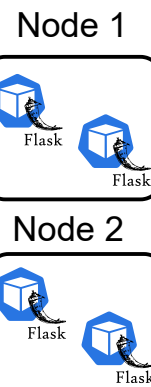


YOLO
1) YOLO Data Retrieval



Gymnasium Environment

Kubernetes Cluster Simulation



Time Series Patching

Vectors

Transformers Core

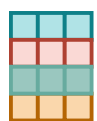
Kubernetes Cluster

Gymnasium Environment

DRL Agent Reward

DRL Action to be implemented

IoT Users



YOLO Features

Cluster State Features

Flask API

Heterogeneous Pods

YOLO Hosted Versions

Gymnasium Flow