Video Prediction Model maps current observation and action sequence to predicted action sequence

Given

Domain randomized RGB images and Depth maps

Task

Learn visual dynamics on domain randomized RGB images and depth maps

Uses large dataset of self-supervised interaction data in the environment

**Advantages**

Generalizable across several tasks and wide variety of environment conditions

**Terminology**

Visuo Spatial Foresight - "identify, integrate, and analyze space and visual form, details, structure and spatial relations"

Self-supervision