



What part of the car
needs to be turned on
when driving at night?



Token Representation



VLM



Continuous Spatial
Relationship



Maintain Rich
Semantic Structure



Decouple Perception
From Reasoning



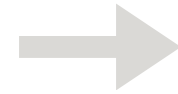
(a) Previous work



What part of the car
needs to be turned on
when driving at night?



Digital Twin
Representation



LLM



Continuous Spatial
Relationship



Maintain Rich
Semantic Structure



Decouple Perception
From Reasoning



(b) DTwinSegger