

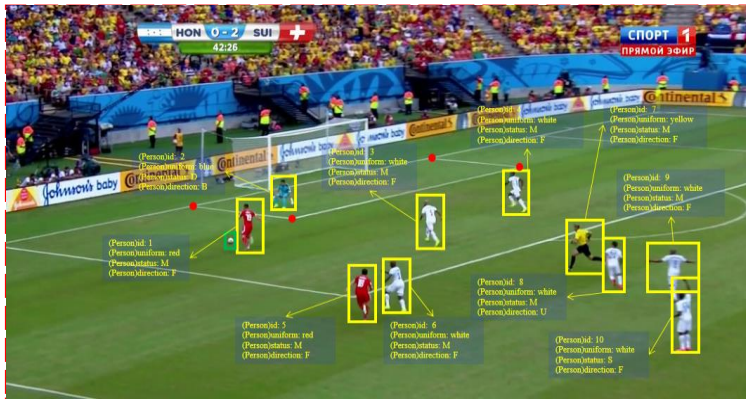
Question

**Semantic
Understanding**

**Visual Task
Selection**

**Required
Information**

“How many players are there in the image?”



Task Modulal

Location Module

Direction Module

Status Module

Color Module

**Key Point
Module**

... ..