

ABSTRACT

AUTONOMOUS DRIVING MODEL BASED ON MULTIMODAL LARGE MODEL, TRAINING METHOD, AND AUTONOMOUS DRIVING METHOD

The present disclosure provides an autonomous driving model based on a multimodal large model, a training method, an autonomous driving method, and relates to the technical field of computers, and in particular to the technical fields of autonomous driving and artificial intelligence. An implementation solution includes: obtaining a training corpus dataset including at least a visual-text aligned corpus and a spatial understanding training corpus for an autonomous driving scenario; encoding, using a vision encoder, visual data in the visual-text aligned corpus to obtain encoded data; mapping the encoded data using a mapping layer; separately processing, using a generation layer, the mapped encoded data, text data, and the spatial understanding training corpus to obtain a first prediction result and a second prediction result of the autonomous driving model; and adjusting a parameter of the autonomous driving model based on at least the first prediction result and the second prediction result. The autonomous driving model trained according to embodiments of the present disclosure has both a multimodal information understanding capability and a reasoning capability in an autonomous driving scenario.

Figure 1