

ABSTRACT

A model processing method and apparatus, a device, and a computer-readable storage medium. The method comprises: ~~obtaining a first computational graph corresponding to a model to be trained and a parallelization strategy of said model, the parallelization strategy of said model comprising at least one of pipeline parallelism, model parallelism, data parallelism, and operator splitting~~ acquiring a first computational graph corresponding to a model to be trained and a parallelization strategy for the model to be trained, wherein the parallelization strategy for the model to be trained includes at least one of the following: pipeline parallelism, model parallelism, data parallelism, and operator split (S601); ~~adding parallelization information in the first computational graph according to the parallelization strategy of said model to obtain a second computational graph~~ adding, according to the parallelization strategy for the model to be trained, parallelization information into the first computational graph, to obtain a second computational graph (S602); ~~determining a distributed computational graph according to the second computational graph and computing resources~~ determining, according to the second computational graph and a computing resource, a distributed computational graph (S603); and ~~training said model according to the distributed computational graph~~ training, according to the distributed computational graph, the model to be trained (S604). Multiple parallelization strategies are supported ~~by~~ based on a graph editing technique of computational graph ~~a computational graph-based graph editing technique~~, so that multiple parallelization strategies can be integrated into a system, thereby realizing a distributed training framework capable of supporting multiple parallelization strategies.