

ABSTRACT

5 The present disclosure provides a neural network operation method and apparatus, an
electronic device and a storage medium. The method includes: determining a network layer to
be processed in a target neural network; determining a target operator and a target chunking
strategy corresponding to the network layer to be processed in the target neural network from a
plurality of determined operators and a plurality of determined chunking strategies, wherein
10 each of the plurality of operators is configured to implement a corresponding function of the
network layer to be processed, and each of the plurality of chunking strategies matches with
operational requirement of a computing device for operating the target neural network; and
operating the target neural network containing the target operator based on the target chunking
strategy corresponding to the network layer to be processed.