

Watermark Information Embedding Method and Apparatus

Abstract

A method and an apparatus for embedding watermark information are disclosed in the present application. The method trains an embedded neural network model using weight information of a target neural network model and target watermark information that is to be embedded into the target neural network model, updates the weight information of the target neural network model according to target watermark embedded data provided by the embedded neural network model, and obtains a target neural network model embedded with the target watermark information. Since the embedded neural network model includes multiple neural network layers, this method increases the complexity of the watermark embedding process, and is able to avoid the problem that watermark information of existing neural network models has poor robustness to watermarking attacks such as overwriting attacks and model compression.

Figure 1