

TRAINING METHODS AND APPARATUSES FOR OBJECT DETECTION SYSTEM

ABSTRACT

Embodiments of this specification provide a training method for an object
5 detection system. The object detection system includes a backbone network and a head
network, the backbone network includes several convolutional layers and several self-
attention layers, and the method includes the following: a training image is input to the
object detection system, where convolution processing is performed on the training
image by using the several convolutional layers, to obtain a convolution representation;
10 self-attention processing is performed based on the convolution representation by using
the several attention layers, to obtain a feature map; and the feature map is processed
by using the head network, to obtain a detection result of a target object in the training
image; a gradient norm of each neural network layer is determined based on object
annotation data and the detection result corresponding to the training image; and for
15 each neural network layer, network parameters of the neural network layer are updated
based on an average of the gradient norms and the gradient norm of the neural network
layer.

20 FIG. 2