# METHODS AND APPARATUSES FOR TRAINING CONTENT UNDERSTANDING MODEL AND CONTENT GENERATION MODEL

## ABSTRACT

Embodiments of this specification relate to a method and an apparatus for training a content understanding model and a content generation model. The method includes: separately training a content understanding model and a content generation model by using an image-text pair formed by an image and a text in a target training set, and then performing sample processing on a noise-containing sample set, including: inputting a first image in any first image-text pair in the noise-containing sample set into the content understanding model to obtain several candidate texts; separately inputting a first text in the first image-text pair and the several candidate texts into the content generation model to obtain multiple candidate images; performing similarity matching between the multiple candidate images and the first image, and determining a target text based on a matching result; and forming a second image-text pair by using the first image and the target text and adding the second image-text pair to the target training set, to continue to train the content understanding model and the content generation model.

FIG. 3