

Integrating Knowledge Distillation of Multiple Strategies

Authors : Min Jindong, Wang Mingxia

Abstract : With the widespread use of artificial intelligence in life, computer vision, especially deep convolutional neural network models, has developed rapidly. With the increase of the complexity of the real visual target detection task and the improvement of the recognition accuracy, the target detection network model is also very large. The huge deep neural network model is not conducive to deployment on edge devices with limited resources, and the timeliness of network model inference is poor. In this paper, knowledge distillation is used to compress the huge and complex deep neural network model, and the knowledge contained in the complex network model is comprehensively transferred to another lightweight network model. Different from traditional knowledge distillation methods, we propose a novel knowledge distillation that incorporates multi-faceted features, called M-KD. In this paper, when training and optimizing the deep neural network model for target detection, the knowledge of the soft target output of the teacher network in knowledge distillation, the relationship between the layers of the teacher network and the feature attention map of the hidden layer of the teacher network are transferred to the student network as all knowledge. in the model. At the same time, we also introduce an intermediate transition layer, that is, an intermediate guidance layer, between the teacher network and the student network to make up for the huge difference between the teacher network and the student network. Finally, this paper adds an exploration module to the traditional knowledge distillation teacher-student network model. The student network model not only inherits the knowledge of the teacher network but also explores some new knowledge and characteristics. Comprehensive experiments in this paper using different distillation parameter configurations across multiple datasets and convolutional neural network models demonstrate that our proposed new network model achieves substantial improvements in speed and accuracy performance.

Keywords : object detection, knowledge distillation, convolutional network, model compression

Conference Title : ICAIIT 2022 : International Conference on Artificial Intelligence and Information Technologies

Conference Location : Boston, United States

Conference Dates : April 21-22, 2022