Introduction

Deep neural network is good, but large, complex and time-consuming, so its application is limited. As a result, we need smaller network; among them is knowledge distillation; (2021-12-29)

Smaller network can achieve same result as larger network, but compared with larger network, smaller network; so we train a teacher, then student mimic teacher; (2021-12-29)

Related Work

Knowledge distillation is proposed by Hinton et al. [12].