

A Gift from Knowledge Distillation: Fast Optimization, Network Minimization, and Transfer Learning

Junmo Kim

Statistical Inference and Information Theory Laboratory, Korea Advanced Institute of Science and Technology, South Korea

E-mail: junmo.kim@kaist.ac.kr

We introduce a novel technique for knowledge transfer, where knowledge from a pretrained deep neural network (DNN) is distilled and transferred to another DNN. While the DNN changes the input space to the output space through many layers sequentially, we define the distilled knowledge to be transferred in terms of flow between layers, which is calculated by computing the inner product between features from two layers. The proposed method of transferring the distilled knowledge as the flow between two layers exhibits three important phenomena: (1) the student DNN that learns the distilled knowledge is optimized much faster than the original model; (2) the student DNN outperforms the DNN that uses the original training procedure; and (3) the student DNN can learn the distilled knowledge from a teacher DNN that is trained at a different task, and the student DNN outperforms a DNN that is trained from scratch.