Self-Supervised Knowledge Transfer via Loosely Supervised Auxiliary Tasks

Seungbum Hong, Jihun Yoon, and Min-Kook Choi
VisionAI, hutom
Seoul, Republic of Korea
{qbration21, jhyoon2020, mkchoi}@hutom.io

Junmo Kim
KAIST
Daejeon, Republic of Korea
junmo.kim@kaist.ac.kr

Abstract

Knowledge transfer using convolutional neural networks (CNNs) can help efficiently train a CNN with fewer parameters or maximize the generalization performance under limited supervision. To enable a more efficient transfer of pretrained knowledge under relaxed conditions, we propose a simple yet powerful knowledge transfer methodology without any restrictions regarding the network structure or dataset used, namely self-supervised knowledge transfer (SSKT), via loosely supervised auxiliary tasks. For this, we devise a training methodology that transfers previously learned knowledge to the current training process as an auxiliary task for the target task through self-supervision using a soft label. The SSKT is independent of the network structure and dataset, and is trained differently from existing knowledge transfer methods; hence, it has an advantage in that the prior knowledge acquired from various tasks can be naturally transferred during the training process to the target task. Furthermore, it can improve the generalization performance on most datasets through the proposed knowledge transfer between different problem domains from multiple source networks. SSKT outperforms the other transfer learning methods (KD, DML, and MAXL) through experiments under various knowledge transfer settings. The source code will be made available to the public[1].

1. Introduction

Knowledge transfer is the most representative training methodology for improving the generalization capability and training efficiency of convolutional neural networks (CNNs). The most widely used knowledge transfer is transfer learning [37, 35], which uses pretrained weights trained on large-scale datasets as initial values for new tasks. Pretrained weights have been used as feature encoders after fine-tuning in different vision tasks such as image classification, object detection, and semantic segmentation [24, 25, 29]. Taskonomy [38, 39] provides an extensive database and analysis approach for studying the effects of transfer learning using pretrained networks on the fine-tuning of target tasks. However, transfer learning using pretrained networks presupposes structural dependencies that must share essentially the same (whole or partial) network structure as the source network. Moreover, questions have been raised on whether transfer learning using pretrained weights can provide appropriate initial weights, which can help in improving convergence or training with less amount of data but not very effective for knowledge transfer to different tasks [12].

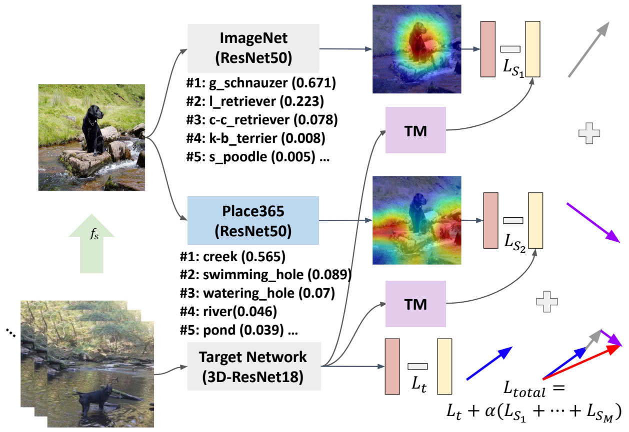

Figure 1. **Motivation of the SSKT.** The CNN responses that can be obtained from the same image vary depending on the type of supervision. SSKT performs auxiliary training using soft labels to convey the previously trained prior knowledge from the source tasks. At this time, the supervision used in the auxiliary training generates a gradient (grey and purple arrows) that can improve the generalization performance of the target task. The gradient to update the weight of the target network is obtained as a linear combination of all losses (red arrow). TM is a transfer module.

Another representative example of knowledge transfer using CNN is knowledge distillation (KD) [14], where pretrained teacher networks distill and deliver dark (hidden) knowledge in the inference output or learned features during student network training. KD methods include loss-based KD, which delivers dark knowledge through loss with a soft

[1] https://github.com/generation21/generation6011