Supplementary Material: Learning to Transfer Learn: Reinforcement Learning-Based Selection for Adaptive Transfer Learning

Linchao Zhu[1,2], Sercan Ö. Arık[1], Yi Yang[2], and Tomas Pfister[1]

[1] Google Cloud AI, Sunnyvale, CA
[2] University of Technology Sydney, Sydney, Australia
{soarik,tpfister}@google.com; {linchao.zhu,yi.yang}@uts.edu.au

Table 1. Top chosen classes from ImageNet source dataset.

Target dataset	Source classes (weights)
Birds	bee eater (1.0); aepyceros melampus (0.95); sea cradle (0.92); barracouta (0.91); valley (0.90); sombrero (0.86); rosehip (0.84); Scottish deerhound (0.82); black swan (0.77); bell pepper (0.77);
Pets	coral reef (0.88); prayer rug (0.88); koala (0.84); fire salamander (0.81) Irish setter (0.79); Arabian camel (0.78); Irish terrier (0.74); leaf beetle (0.72); Rottweiler (0.71); giant schnauzer (0.70);
Cars	desktop computer (0.80); butternut squash (0.76); barrel, cask (0.65); weevil (0.60); pool table (0.56); clumber (0.54); passenger car (0.50); race car (0.49); washer (0.46); terrapin (0.45); seat belt (0.32);
Aircraft	bagel, beigel (0.97); ballplayer, baseball player (0.84); freight car (0.80); teapot (0.83); crate (0.78); velvet (0.74); electric locomotive (0.68); pirate, pirate ship (0.45); amphibious vehicle (0.28); airliner (0.10);
Food	caldron, cauldron (0.84); menu (0.81); seashore, coast, seacoast (0.74) acorn squash (0.73); dining table, board (0.67); globe artichoke (0.67); mushroom (0.65); chocolate sauce, chocolate syrup (0.62); plate (0.54);

1 High-weight classes from ImageNet

For the target datasets, i.e., Birdsnap, Oxford-IIIT Pets, Stanford Cars, FGVC Aircraft, Food-101, we list the high-weight classes from the source ImageNet in Table 1. These classes are most related to the target dataset. Some top classes do not from the same categories, but they are visually similar. For instance, the top related class for "Stanford Cars" is "desktop computer", where the "desktop computer" class like boxes. The "valley" class matches the background in some examples in the Birdsnap dataset.

For the Aircraft dataset, the "teapot" class is assigned with weight 0.83. We observe that the spout of teapot and teapot handle resemble the wing of an airplane. This indicates that our model is able to learn features that are implicitly relevant to the target samples and would benefit the learning process, even though the features may belong to different classes. Our L2TL yields ranking of the source data samples, which can be utilized as new forms of interpretable insights for model developers.

2 More details on REINFORCE

We add more details about the REINFORCE algorithm. The objective is $J(\mathbf{\Phi}) = E_{\lambda \sim \pi_{\mathbf{\Phi}}} r(\lambda)$, where $\mathbf{\Phi}$ is the parameters to be optimized, $r(\lambda)$ is the reward, and λ is a sampled action. We aim to optimize parameters by maximizing $J(\mathbf{\Phi})$. The gradient of $J(\mathbf{\Phi})$ is

$$\nabla_{\mathbf{\Phi}} J(\mathbf{\Phi}) = E_{\lambda \sim \pi_{\mathbf{\Phi}}(\lambda)} \nabla_{\mathbf{\Phi}} \log \pi_{\mathbf{\Phi}}(\lambda)) r(\lambda). \tag{1}$$

Monte Carlo sampling is used to obtain N sequences to approximate the policy gradients:

$$\nabla_{\mathbf{\Phi}} J(\mathbf{\Phi}) \approx \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \nabla_{\mathbf{\Phi}} \log \pi_{\mathbf{\Phi}}(\lambda_{i,t}) r(\lambda_{i,t}), \tag{2}$$

where T is the length of an episode.