

作者：小哲

微信公众号：小哲AI

原文来自微信公众号

github: <https://github.com/lxztju>

自监督学习修炼之SimCLRv2

1. 论文摘要

利用少量有标注样本及大量无标注样本进行学习的一个范式是无监督预训练然后进行有监督的微调。尽管与半监督学习的方法相比我们的方法是一种任务无关(task-agnostic)的方式，但是实际的实验已经证明这是一种非常高效的方案。我们方法中的一个主要的部分就是在预训练与微调的时候使用一个大的（宽度和深度）神经网络。我们发现，有标注的数据集的数目越少，那么我们的方案就能从大网络中获得更大的提升，在微调之后，大网络keying获得更大程度的提升，并且就可以在准确率损失很小的情况下蒸馏给小的网络，蒸馏过程中使用无标注数据集，但是这次是一个确定任务(task-specific)的方式。提出的半监督学习算法可以总结为以下三个步骤：**1. 利用大的resnet进行无监督的预训练， 2. 在少量有标注数据集上有监督的微调， 3. 利用无标注蒸馏提炼并转换特定任务的知识。**利用resnet50和1%的有标注样本（每类小于13张），这个过程能够实现73.9%的ImageNet top1的准确率。与过去的最优的算法相比在label的高效利用上可以获得10倍的提升。利用10%的有标注样本，利用我们的方案resnet50实现了77.5%的top1准确率超出利用所有样本的标准监督训练。

总结起来：

1. 首先利用大量无标注样本进行无监督预训练（大网络resnet152）
2. 然后利用有标注样本进行微调
3. 利用无标注样本集将大网络学到的知识蒸馏给小网络（resnet50）

2. 论文主要思想

相较于SimCLR算法，这里的v2版本增加的主要就是采用蒸馏的思想先在大网络上进行训练然后将知识蒸馏给小网络。

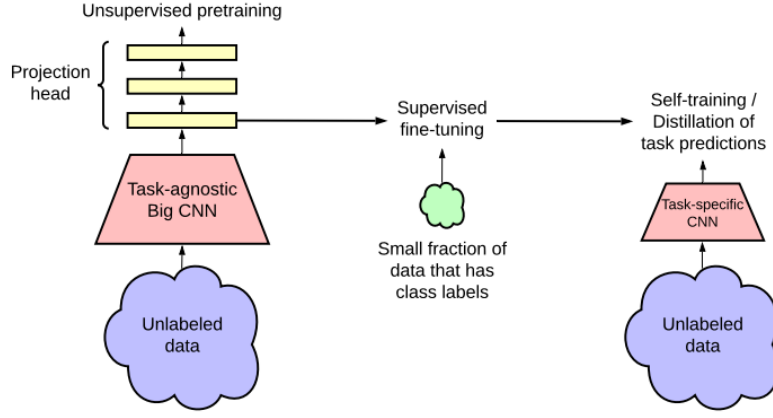


Figure 3: The proposed semi-supervised learning framework leverages unlabeled data in two ways: (1) task-agnostic use in unsupervised pretraining, and (2) task-specific use in self-training / distillation.

无监督预训练的loss函数与SimCLR的loss function一致。

$$\ell_{i,j}^{\text{NT-Xent}} = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)},$$

利用无标注数据集进行蒸馏所用的loss:

$$\mathcal{L}^{\text{distill}} = - \sum_{\mathbf{x}_i \in \mathcal{D}} \left[\sum_y P^T(y|\mathbf{x}_i; \tau) \log P^S(y|\mathbf{x}_i; \tau) \right]$$

论文中蒸馏所使用的数据集是无标注的数据集，也可以使用有标注样本，如果使用有标注样本在考虑教师网络输出的同时也要考虑真实label, ground-truth的值，将二者进行加权作为最终的损失函数。

$$\mathcal{L} = -(1 - \alpha) \sum_{(\mathbf{x}_i, y_i) \in \mathcal{D}^L} \left[\log P^S(y_i|\mathbf{x}_i) \right] - \alpha \sum_{\mathbf{x}_i \in \mathcal{D}} \left[\sum_y P^T(y|\mathbf{x}_i; \tau) \log P^S(y|\mathbf{x}_i; \tau) \right].$$

论文地址: [Big Self-Supervised Models are Strong Semi-Supervised Learners](#)