# 浏览式阅读

## 1 自己的总结、评价以及应用

本文就介绍了这样一个东西：Deep mutual learning（用小网络实现大网络的功能，且功耗更小）

In this paper, we present a deep mutual learning (DML) strategy where, rather than one way transfer between a static pre-defined teacher and a student, an ensemble of students learn collaboratively and teach each other throughout the training process.

翻译：

在本文中，我们提出了一种深度相互学习（DML）策略，其中，不是在静态的预定义老师和学生之间进行单向转移，而是在整个培训过程中，一群学生会积极地学习协作并互相教书。

独特之处：

Surprisingly, it is revealed that no prior powerful teacher network is necessary（不需要强网络） – mutual learning of a collection of simple student networks works（只需要简单网络即可）, and moreover outperforms distillation from a more powerful yet static teacher.

## 2 文章的主要问题（abstract、疑问句中）

本文就介绍了这样一个东西：Torchreid以及它的几大特性

a software library built on PyTorch that allows fast development andend-to-end training and evaluation of deep re-ID models

## 3 结论（abstract以及conclusion中）

## 4 思路脉络（小标题中的关键句）

1. **Introduction**

深度学习存在的drawback：

This has the drawback that they may be slow to execute or demand large memory to store, limiting their use in applications or platforms with low memory or fast execution requirements.

目前已有的解决方法：

Achieving compact yet accurate models has been approached in a variety of ways including explicit frugal architecture design [8], model compression [20], pruning [13], binarisation [18] and most interestingly model distillation [7].

Distillation-based model compression：优点与缺点

small networks often have the same representation capacity as large networks; but compared to large networks they are simply harder to train and find the right parameters that realise the desired function.

如何解决问题：

To better learn a small network, the distillation approach starts with a powerful (deep and/or wide) teacher network (or network ensemble), and then trains a smaller student network to mimic the teacher[7, 2, 16, 3].

我们做的工作：

In this paper we explore a different but related idea to model distillation – that of mutual learning

本文就干了这一件事：（a peer-teaching based scenario　mutual learning）（关键：collaborative learning by small peers）

This motivates us to design a generic framework that provides a standardised data-loading interface, basic training pipelines compatible with different re-ID models, and more importantly, is easy to extend.

in mutual learning we start with a pool of untrained students who learn simultaneously to solve the task together. Specifically, each student is trained with two losses: a conventional supervised learning loss, and a mimicry loss that aligns each student’s class posterior with the class probabilities of other students.

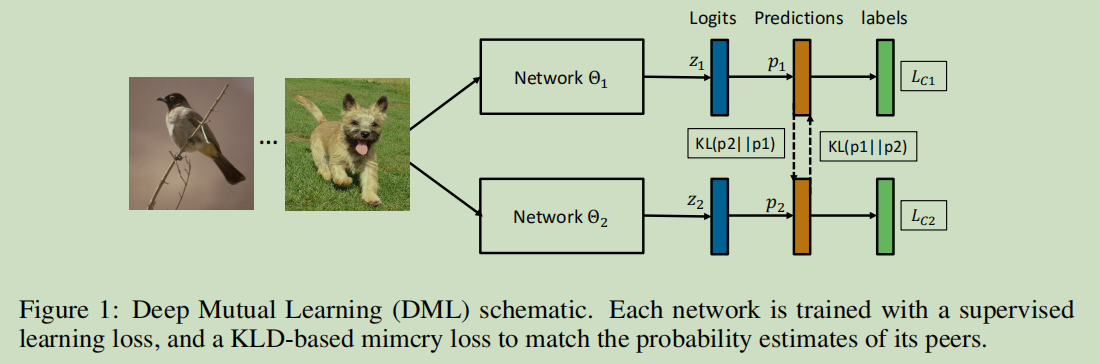
Here we address dispensing with the teacher altogether, and allowing an ensemble of students to teach each other in mutual distillation.

最近的一些变化：

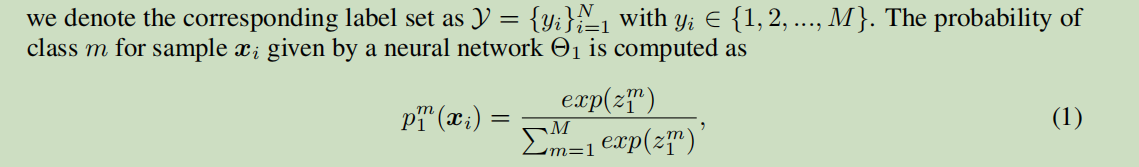
For CNN model learning, Torchreid currently implements two training pipelines, which are classification with softmax2 loss and metric learning with triplet3 loss, the two widely used (and most effective) objective functions in the literature.

2 Deep Mutual Learning3 Main Modules

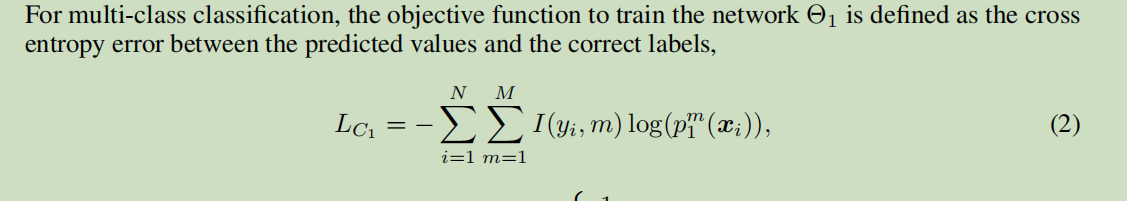
2.1 Formulation



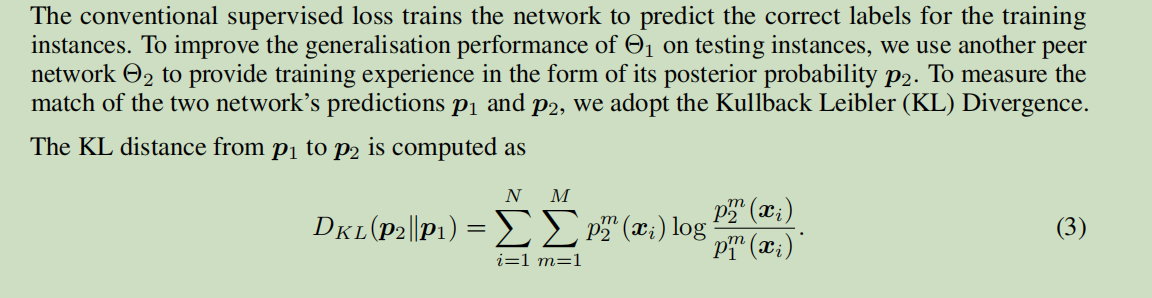
每个网络测出的概率



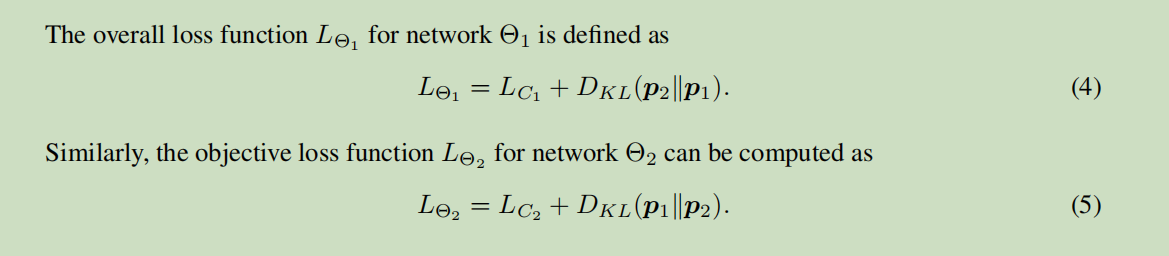
每个网络的目标（损失）函数



两个peer network预测结果的KL distance

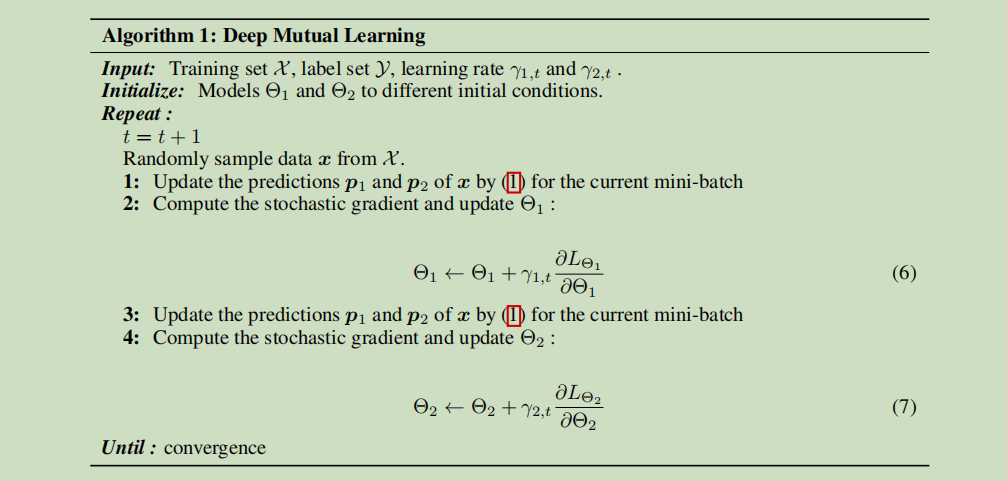


然后，新的目标（损失）函数就变为：



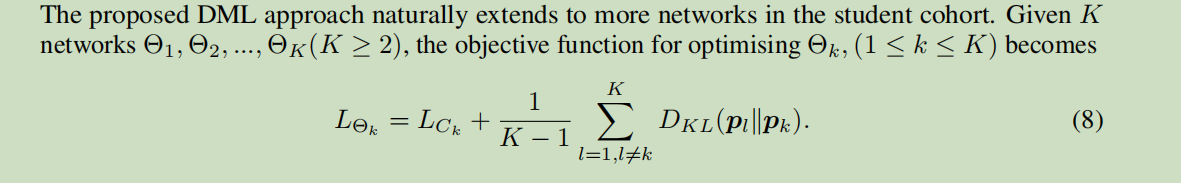
2.2 Optimisation

优化算法如下：



2.3 Extension to Larger Student Cohorts（本质上讲就是：1个做学生，剩下k-1个做老师）

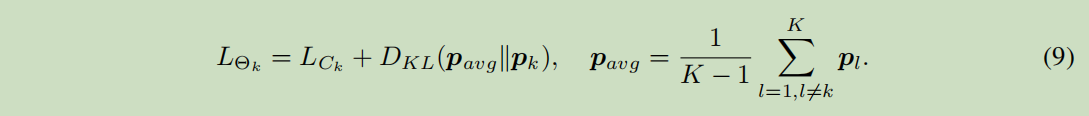
K个network时每一个network的目标（损失）函数为：



Equation (8) indicates that with K networks, DML for each student effectively takes the other K − 1 networks in the cohort as K −1 teachers to provide learning experience.

再次优化：

With more than two networks, an interesting alternative learning strategy for DML is to take the ensemble of all the other K − 1 networks as a single teacher to provide an averaged learning experience, which would be very similar to the distillation approach but performed at each mini-batch model update.



3.1 Data

模型搭建过程中最难的一步：

the construction of unified data loaders for preparing data

第一个部分Data\_loader

The training and test data loaders are wrapped in a high-level class called DataManager,

DataManager包括这些工作：sampling strategy, data augmentation methods and data loaders.有两种分类：ImageDataManager 和 VideoDataManager

3.2 Engine

Engine作为一个module可以provide universal training loops and other reusable features, such as data parsing, model checkpointing and performance measurement，同时，它还提供了两种学习范式learning paradigms

分别是：

1 classification with softmax loss (ImageSoftmaxEngine & VideoSoftmaxEngine)

2 metric learning with triplet loss (ImageTripletEngine & VideoTripletEngine)

另外，torchreid还提供了一些training tricks（训练技巧）：

1 to reduce overfitting，the label smoothing regulariser (Szegedy et al., 2016) is implemented for the softmax pipeline.

2 for better transfer learning，the pipeline allows the pre-trained CNN layers to be frozen during early training (Geng et al., 2016) where the layers are specified by users.

Visualisation toolkit

Torchreid为我们提供了一些函数：

1 visrank,which can visualise the ranking result of a re-ID CNN by saving for each query image the top-k similar gallery images (k is decided by users)将gallery images中排名top-k的照片展示出来

2 visactmap,, which stands for visualising activation maps. Given an input image, the activation map can be used to analyse where the CNN focuses on to extract features

3.3 Models

The currently available models are listed below,

• ImageNet classification models: ResNet (He et al., 2016), ResNeXt (Xie et al., 2017),

SENet (Hu et al., 2018), DenseNet (Huang et al., 2017), Inception-ResNet-V2 (Szegedy

et al., 2017), Inception-V4 (Szegedy et al., 2017), and Xception (Chollet, 2017).

• Lightweight models: NASNet (Zoph et al., 2018), MobileNetV2 (Sandler et al., 2018),

ShuffleNet(V2) (Zhang et al., 2018; Ma et al., 2018), and SqueezeNet (Iandola et al., 2016).

• Re-ID specific models: MuDeep (Qian et al., 2017), ResNet-mid (Yu et al., 2017),

HACNN (Li et al., 2018), PCB (Sun et al., 2018), MLFN (Chang et al., 2018), OSNet (Zhou

et al., 2019b), and OSNet-AIN (Zhou et al., 2019a)

**4 Discussion**

## 5 难理解点