

Chapter 2-2

PyTorch 及 Dataloader 介绍

The Introduction of Pytorch and Dataloader

分享人

Dalian Maritime University

臧景奇

ICDC department, Dianhang Association

jonathan@dlmu.edu.cn

第一部分

Common Environment of Deep Learning

深度学习的常见生态

✓ TensorFlow ✓ Keras ✓ PyTorch

TensorFlow 简介

TensorFlow是一个软件库或框架，由Google团队设计，以最简单的方式实现机器学习和深度学习概念，通常用于工业、企业的机器学习开发。

其缺点也很明显，TensorFlow一代与二代不兼容，而大量的工作和成熟代码都是基于TensorFlow 1.x 版本搭建的，这导致许多旧工程无法运行。

官网: <https://tensorflow.google.cn/>

Keras 简介

Keras 是一个用 Python 编写的高级神经网络 API，简便易于理解，同时支持卷积神经网络(CNN)和循环神经网络(RNN)，可以在 CPU 和 GPU 上无缝运行。

但是由于 Keras 本身是一个中间层，它是通过调用 TensorFlow 或 theano 实现的功能，因此如果需要训练大尺寸模型的话可能会频繁遇到 GPU 内存溢出的问题，且速度较慢。

官网: <https://keras.io/>

PyTorch 简介



- PyTorch 是一个开源的 Python 机器学习库，底层由 C++ 实现，应用于人工智能领域
- 它主要由 Facebook 的人工智能研究团队开发，被广泛应用于学术研究界。

优点

- ✓ **上手快**：掌握 Numpy 和基本深度学习概念即可上手
- ✓ **代码简洁灵活**：用 `nn.module` 封装使网络搭建更方便，基于动态图机制，更灵活
- ✓ **Debug方便**：调试 PyTorch 就像调试 Python 代码一样简单
- ✓ **开发者多**：GitHub 上贡献者 (Contributors) 已达 1100+
- ✓ **兼容性好**：不会像 TensorFlow 那样版本之间存在自我冲突现象
- ✓ **运行速度快**：相较于 Keras, PyTorch 的性能要好得多

官网: <https://pytorch.org/>

PyTorch 简介



- 使用PyTorch，可以轻松地解决线性回归、拟合、分类等问题，并且具有良好的可视化性。

应用

- ✓ 使用 PyTorch，我们可以实现将蚂蚁和蜜蜂的图片进行区分。不仅仅是蚂蚁蜜蜂，各种特征鲜明的动植物、物体等都可用PyTorch进行区分。



- ✓ 我们还可以根据黄金的历史价格，来预测未来黄金价格的走向。不仅仅是黄金，任何股票或比特币等资产都可被用于预测中。

官网: <https://pytorch.org/>

第二部分

Installation Demo of PyTorch-CPU Environment

PyTorch-CPU 环境的安装演示

有安装 GPU 环境需求的可申请单独辅导

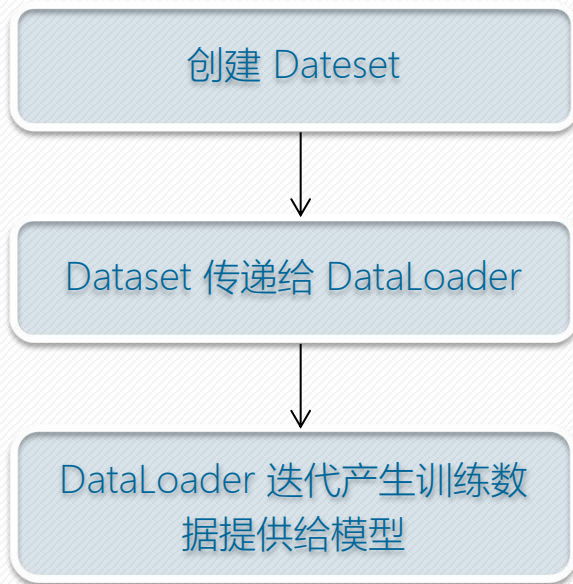
第三部分

Introduction of DataLoader DataLoader 简介

有安装 GPU 环境需求的可申请单独辅导

DataLoader 与 Dataset 简介

- 机器学习最重要部分之一就是数据集，只有强大的数据集才能训练出优秀的机器学习模型
- 数据集通常规模庞大，例如 CIFAR10 数据集中就包含了 60000 张彩色图像
- 自己寻找数据集的工程量是非常大的，所以 PyTorch 很贴心地为我们提供了大量的优秀数据集使用，并将其封装为 **Dataset** 类，可直接调用



数据集的传递流程

常用数据集介绍

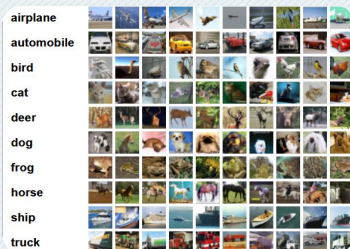
FASHION MNIST

Fashion-MNIST 是一个包含了来自10种类别的共70000个不同商品的正面灰度图片的数据集，每张图片的大小为 28×28 ，训练集/验证集的划分为：60000/10000



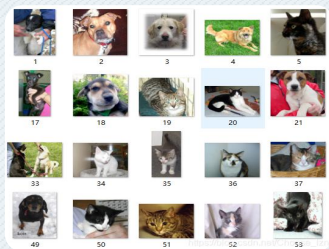
CIFAR-10

CIFAR-10 是一个包含了来自10种不同类别的共60000张彩色图片的数据集，每张图片的大小为 32×32 ，训练集/验证集的划分为：50000/10000



CATS vs. DOGS

Cats vs. Dogs 是一个包含了猫与狗各12500张彩色图片的数据集，数据集中的图片尺寸都不尽相同，没有进行统一的裁剪。



THANKS!

分享人

Dalian Maritime University

臧景奇

ICDC department, Dianhang Association

jonathan@dlmu.edu.cn