

CANN 训练营

新手班模型开发课学员手册

文档版本

01

发布日期

2022-02-24



版权所有 © 华为技术有限公司 2022。 保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI 和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编： 518129

网址： <https://www.huawei.com>

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

目 录

1 课程目标.....	1
2 AI 模型开发基础知识入门.....	2
2.1 具备 Python 编程经验.....	2
2.2 了解深度学习和神经网络.....	2
2.3 了解 TensorFlow AI 框架.....	3
2.4 昇腾 AI 全栈架构.....	3
2.5 异构计算架构 CANN.....	4
2.6 了解基于 CANN 的模型开发流程.....	6
2.7 了解遇到问题如何求助.....	6
2.8 小作业 1.....	6
3 TensorFlow AI 模型迁移详解.....	8
3.1 本章学习目标.....	8
3.2 为什么要做模型迁移.....	8
3.3 了解两种模型迁移方式.....	9
3.4 TensorFlow AI 模型自动迁移详解.....	9
3.5 TensorFlow AI 模型手工迁移详解.....	9
3.6 小作业 2.....	9
4 TensorFlow AI 模型训练详解.....	11
4.1 本章学习目标.....	11
4.2 学习资源.....	11
4.3 小作业 3.....	11
5 AI 模型迁移和训练实战演练.....	13
5.1 大作业.....	13
5.2 附加题.....	14
6 附录：作业提交格式.....	15

1 课程目标

欢迎大家来到《CANN 训练营》新手班模型开发课，通过本次学习，您将收获以下知识：

- 了解异构计算架构 CANN 在神经网络训练中发挥的主要作用。
- 掌握如何基于 CANN 将 TensorFlow 模型迁移到昇腾 AI 处理器上。
- 掌握如何在昇腾 AI 处理器上进行模型训练，感受昇腾 AI 的极致性能。
- 掌握如何查看训练日志和训练结果，具备基本的问题定界、定位能力。

2 AI 模型开发基础知识入门

- 2.1 具备 Python 编程经验
- 2.2 了解深度学习和神经网络
- 2.3 了解 TensorFlow AI 框架
- 2.4 昇腾 AI 全栈架构
- 2.5 异构计算架构 CANN
- 2.6 了解基于 CANN 的模型开发流程
- 2.7 了解遇到问题如何求助
- 2.8 小作业 1

2.1 具备 Python 编程经验

观看学习视频，请点击：[LINK](#)

本课程中的示例代码、练习涉及 Python 语言的如下基础知识，建议您在学习本课程前先学习这部分内容：

- 使用位置和关键字参数定义和调用函数
- 字典、列表、集合（创建、访问和迭代）
- for 循环，for 具有多个迭代器变量的循环（例如，for a, b in [(1,2), (3,4)]）
- if/else 条件块和条件表达式
- 字符串格式（例如，'%.2f' % 3.14）
- 变量、赋值、基本数据类型（int, float, bool, str 等）

2.2 了解深度学习和神经网络

观看学习视频，请点击：[LINK](#)

为了让计算机掌握人类理解的知识，需要构筑一个由简单概念组成的多层连接网络来定义复杂对象，计算机通过对这个网络的迭代计算与训练后，可以掌握这个对象的特征，一般称这种方法为深度学习（DeepLearning，DL）。

互联网的发展产生了庞大的数据量，为深度学习的发展提供了更大的机会，也让人工智能成为当今的热点，而深度神经网络成了热点中的热点。深度学习经过一系列的发展之后，展现出巨大的应用价值，不断受到工业界、学术界的密切关注。深度学习在图像、语音、自然语言处理、大数据特征提取和广告点击率预估方面取得明显进展。

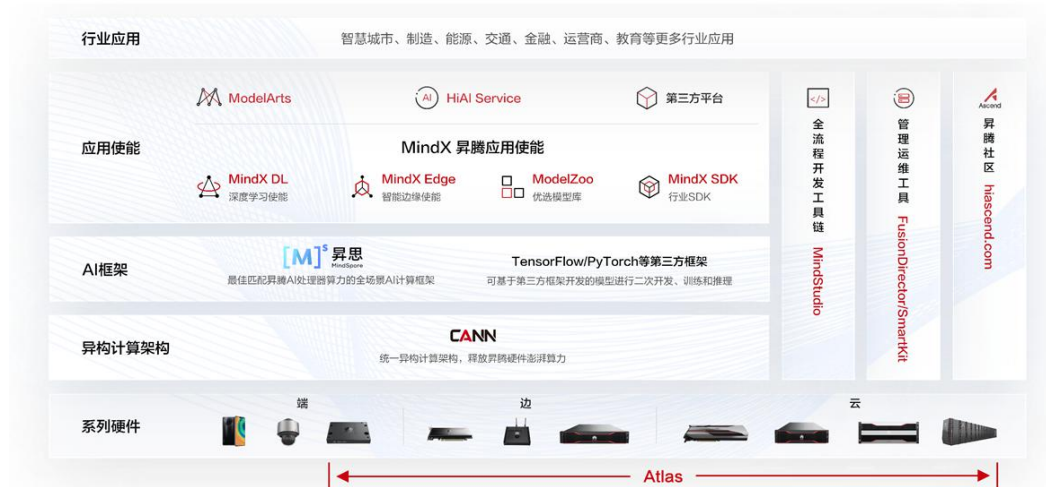
关于深度学习的基本原理、基本概念等介绍，请参见：[Link](#)。

2.3 了解 TensorFlow AI 框架

本课程用于将基于 TensorFlow 的 Python API 开发的训练脚本，将模型迁移到昇腾 AI 处理器上执行训练，因此需要了解 TensorFlow AI 框架。具体请参考 <https://www.tensorflow.org/>。

2.4 昇腾 AI 全栈架构

观看学习视频，请点击：[LINK](#)。



昇腾 AI 全栈可以分成四个大部分：

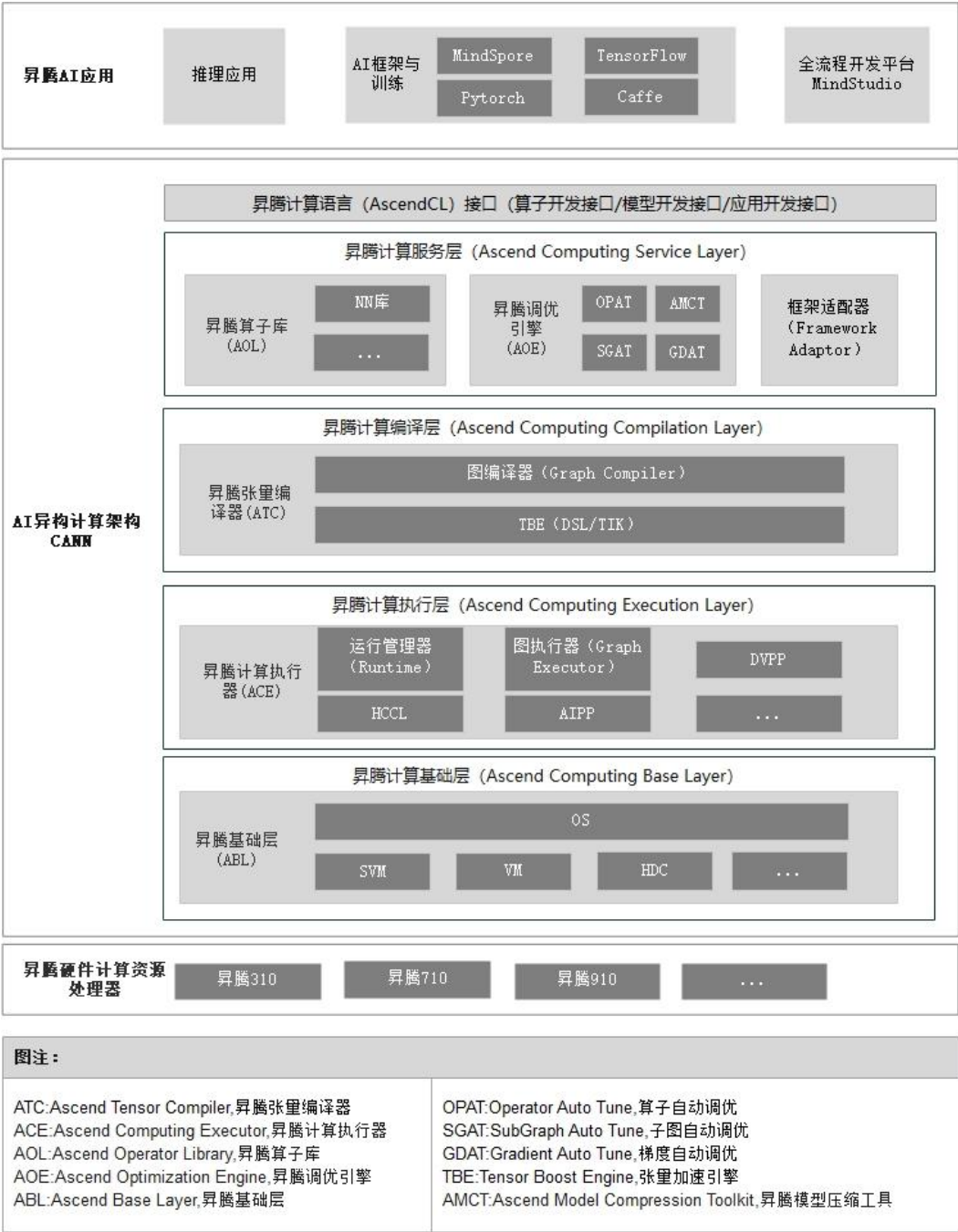
1. 应用使能层面，此层面通常包含用于部署模型的软硬件，例如 API、SDK、部署平台，模型库等等。
2. AI 框架层面，此层面包含用于构建模型的训练框架，例如昇思 MindSpore、TensorFlow、Pytorch 等。
3. 异构计算架构，偏底层、偏通用的计算框架，用于针对上层 AI 框架的调用进行加速，力求向上支持多种 AI 框架，并在硬件上进行加速。
4. 计算硬件，本层是 AI 计算的底座，有了强力的芯片及硬件设备，上层的加速才有实施的基础。

2.5 异构计算架构 CANN

观看学习视频，请点击：[LINK](#)。

华为公司面向计算机视觉、自然语言处理、推荐系统、类机器人等领域量身打造了基于“达芬奇（DaVinci）架构”的昇腾（Ascend）AI 处理器，开启了智能之旅。为提升用户开发效率和释放昇腾 AI 处理器澎湃算力，同步推出针对 AI 场景的异构计算架构 CANN（Compute Architecture for Neural Networks），CANN 通过提供多层次的编程接口，以全场景、低门槛、高性能的优势，支持用户快速构建基于 Ascend 平台的 AI 应用和业务。

昇腾 AI 异构计算架构（Compute Architecture for Neural Networks，CANN）被抽象成五层架构，如下图所示。



1. 昇腾计算语言接口

昇腾计算语言 (Ascend Computing Language, AscendCL) 接口是昇腾计算开放编程框架，是对低层昇腾计算服务接口的封装。它提供 Device (设备) 管理、Context (上下文) 管理、Stream (流) 管理、内存管理、模型加载与执行、算子加载与执行、媒体数据处理、Graph (图) 管理等 API 库，供用户开发人工智能应用调用。

2. 昇腾计算服务层

本层主要提供昇腾计算库，例如神经网络 (Neural Network, NN) 库、线性代数计算库 (Basic Linear Algebra Subprograms, BLAS) 等；昇腾计算调优引擎库，例如算子调优、子图调优、梯度调优、模型压缩以及 AI 框架适配器。

3. 昇腾计算编译引擎

本层主要提供图编译器（Graph Compiler）和 TBE（Tensor Boost Engine）算子开发支持。前者将用户输入中间表达（Intermediate Representation, IR）的计算图编译成 NPU 运行的模型。后者提供用户开发自定义算子所需的工具。

4. 昇腾计算执行引擎

本层负责模型和算子的执行，提供如运行时（Runtime）库（执行内存分配、模型管理、数据收发等）、图执行器（Graph Executor）、数字视觉预处理（Digital Vision Pre-Processing, DVPP）、人工智能预处理（Artificial Intelligence Pre-Processing, AIPP）、华为集合通信库（Huawei Collective Communication Library, HCCL）等功能单元。

5. 昇腾计算基础层

本层主要为其上各层提供基础服务，如共享虚拟内存（Shared Virtual Memory, SVM）、设备虚拟化（Virtual Machine, VM）、主机-设备通信（Host Device Communication, HDC）等。

2.6 了解基于 CANN 的模型开发流程

观看学习视频，请点击：[LINK](#)

昇腾 910 AI 处理器是华为在 2019 年发布的人工智能（AI）专用的神经网络处理器。当前业界大多数训练脚本基于 TensorFlow 的 Python API 开发，默认运行在 CPU/GPU/TPU 上，为了使其能够利用昇腾 910 AI 处理器的澎湃算力执行训练，提升训练性能，我们需要基于 CANN 对 TensorFlow 模型进行适配和训练，以下是基于 CANN 的模型开发流程：[LINK](#)

2.7 了解遇到问题如何求助

模型调试和调优过程中，不知道该怎么解决，或者有优化建议要反馈，您可以借助以下渠道：

- 在华为方集中组织培训的场景下，可求助培训或对应课程的接口人。
- 开发者自学的场景下：
 - 可在 [ModelZoo](#) 仓上提 Issue，该仓的接口人会定期处理问题。
 - 可在 [昇腾社区论坛](#) 中查阅经验贴、或者发帖，论坛的接口人会定期处理问题。

2.8 小作业 1

1. （多选题）昇腾目前支持哪些 AI 深度学习框架？【分值：2 分】
 - A. 昇思 MindSpore
 - B. TensorFlow
 - C. Pytorch

2. （多选题）基于 CANN 的模型开发需要经历哪些环节？【分值：2 分】
 - A. 模型迁移
 - B. 模型训练
 - C. 模型调优
3. （单选）TensorFlow 模型迁移过程主要依赖 CANN 软件栈的哪块能力？【分值：2 分】
 - A. 昇腾调优引擎 AOE
 - B. 昇腾计算执行引擎 ACE
 - C. 框架适配器 Framework Adaptor
4. （多选题）模型调试过程中，遇到问题如何求助？【分值：2 分】
 - A. 到 [ModelZoo](#) 仓上提 Issue，该仓的接口人会定期处理问题。
 - B. 在[昇腾社区论坛](#)中查阅经验贴、或者发帖，论坛的接口人会定期处理问题。
 - C. 在华为方集中组织培训的场景下，可求助培训或对应课程的接口人。

3 TensorFlow AI 模型迁移详解

- 3.1 本章学习目标
- 3.2 为什么要做模型迁移
- 3.3 了解两种模型迁移方式
- 3.4 TensorFlow AI 模型自动迁移详解
- 3.5 TensorFlow AI 模型手工迁移详解
- 3.6 小作业 2

3.1 本章学习目标

通过本节课程，您将学到：

- 了解为什么要做模型迁移
- 了解模型迁移的两种迁移方式
- 掌握如何进行模型迁移

3.2 为什么要做模型迁移

观看学习视频，请点击：[LINK](#)

目前人工智能领域内，AI 算法模型搭建方面的技艺已经是炉火纯青，市面上用于 AI 模型搭建的深度学习框架，除了华为开源的昇思 MindSpore，还有 Google 的 TensorFlow、Facebook 的 PyTorch、Caffe 等。

但是 AI 模型训练却是一个漫长的过程，随着模型参数量、样本的增多，训练一个模型动辄以月为单位，此时一款高性能 AI 处理器简直是雪中送炭。昇腾 910 AI 处理器是华为在 2019 年发布的人工智能（AI）专用的神经网络处理器，其算力高达 256T，最新款算力高达 310T，是业界主流处理器算力的 2 倍，基于昇腾 910 AI 处理器的昇腾系列硬件，比如 Atlas 800 训练服务器，算力更是大到惊人，此时你一定跃跃欲试了。

除了昇思 MindSpore 外，TensorFlow 等其他深度学习框架下的模型并不能直接在昇腾 910 AI 处理器上训练，为了使其充分利用昇腾 910 AI 处理器的澎湃算力来提升训练性能，我们需要借助异构计算架构 CANN 的 Plugin 适配层转换，使转换后的模型能够高效运行在昇腾 910 AI 处理器上。值得庆幸的是，目前，CANN 已经能够支持多种主流 AI 框架，包括昇思 MindSpore、TensorFlow、PyTorch、飞浆、ONNX 等，并且开发者只需要非常少的改动，即可快速搞定算法移植，大大减少切换平台的代价。

下面的课程中，我们主要以 TensorFlow 模型迁移和训练为例进行介绍。

3.3 了解两种模型迁移方式

观看学习视频，请点击：[LINK](#)

将 TensorFlow 网络模型迁移到昇腾 AI 处理器执行训练，主要有两种方式：

- 一种是自动迁移方式。通过迁移工具对原始脚本进行 AST 语法树扫描，可自动分析原生的 TensorFlow API 在昇腾 AI 处理器上的支持度，并将原始的 TensorFlow 训练脚本自动迁移成昇腾 AI 处理器支持的脚本，对于无法自动迁移的 API，可以参考工具输出的迁移报告，对训练脚本进行相应的适配修改。
- 一种是手工迁移方式。算法工程师需要人工分析 TensorFlow 训练脚本中的 API 支持度情况，并且参照文档逐一手工修改不支持的 API，以便在昇腾 AI 处理器上训练，该种方式较为复杂，我们建议大家优先使用自动迁移方式。

3.4 TensorFlow AI 模型自动迁移详解

观看学习视频，请点击：[LINK](#)

相关学习文档，请点击：[LINK](#)

3.5 TensorFlow AI 模型手工迁移详解

视频：[LINK](#)

课程：[LINK](#)（请观看第一章内容，可不关注 ModelArts 训练部分的介绍，后续课程会有讲解）

在线实验：[LINK](#)

文档：[LINK](#)

3.6 小作业 2

1. （多选题）目前自动迁移工具支持自动迁移哪些 API？【分值：2 分】
 - A. TensorFlow 1.15 官方 API
 - B. Horovod 官方 API

- C. 原生 Keras API
- D. CuPy API
- 2. （多选题）模型迁移之前需要做哪些准备工作？【分值：2 分】
 - A. 在 GPU 或 CPU 进行训练复现，训练跑通
 - B. 在 GPU 或 CPU 进行训练复现，Loss 收敛
 - C. 在 NPU 进行训练复现，训练跑通
 - D. 在 NPU 进行训练复现，Loss 收敛
- 3. （多选题）当前昇腾 910 AI 处理器上支持迁移 TensorFlow 的哪几类 API 开发的训练脚本？【分值：2 分】
 - A. Estimator
 - B. Sess.run
 - C. Keras

4 TensorFlow AI 模型训练详解

4.1 本章学习目标

4.2 学习资源

4.3 小作业 3

4.1 本章学习目标

通过本节课程，您将学习到：

- 了解 ModelArts 训练之前各项准备工作
- 了解 OBS、ModelArts 以及 Pycharm 插件关系
- 掌握如何使用 ModelArts 平台进行网络训练
- 掌握如何查看训练日志和训练结果，具备基本的问题定界、定位能力。

4.2 学习资源

观看学习视频，请点击：[LINK](#)

相关学习文档，请点击：[LINK](#)

4.3 小作业 3

1. （多选题）在 ModelArts 训练之前，需要做哪些准备工作？【分值：2 分】
 - A. 注册华为云账号，申请华为云代金券
 - B. 获取数据集和训练脚本并上传到 OBS
 - C. 安装 Pycharm IDE 及 Toolkit 插件
 - D. 对 TensorFlow 训练代码进行适配
 - E. 在 Pycharm 创建并下发训练任务

2. （多选题）为在 ModelArts 训练，需要对 TensorFlow 训练代码进行哪些方面的适配？【分值：2 分】
 - A. 需要在训练脚本中调用 moxing 接口，将 OBS 数据集传输至 ModelArts
 - B. 需要在训练脚本中调用 moxing 接口，将 ModelArts 的训练结果传输回 OBS
 - C. 需要在训练脚本中指定训练性能数据保存路径
 - D. 需要在训练脚本中指定各层算子计算结果的 dump 数据保存路径
3. （多选题）在 ModelArts 训练结束后，可以到哪里查看训练结果文件和训练日志？【分值：2 分】
 - A. Pycharm IDE 界面
 - B. OBS
 - C. PC 端本地目录

5

AI 模型迁移和训练实战演练

5.1 大作业

5.2 附加题

5.1 大作业

1. 使用 Tensorflow1.15 实现 LeNet 网络的 mnist 手写数字识别。硬件平台不限，可以基于 window 或者 linux 系统，尽量给出整个过程的截图，并在最后给出 Loss 或者 accuracy 运行结果。参考链接 [Gitee](#) 或 [Github](#)。【分值：10 分】

步骤 xx 截图 xx

步骤 xx 截图 xx

```
step 1100, training acc 0.54 time: 4.022 (ms)
step 1200, training acc 0.6 time: 3.01 (ms)
step 1300, training acc 0.5 time: 2.006 (ms)
step 1400, training acc 0.64 time: 2.005 (ms)
step 1500, training acc 0.62 time: 3.009 (ms)
step 1600, training acc 0.54 time: 2.006 (ms)
step 1700, training acc 0.58 time: 2.005 (ms)
step 1800, training acc 0.72 time: 5.517 (ms)
step 1900, training acc 0.7 time: 3.007 (ms)
PS D:\gitspace\sample-lenet>
```

2. 采用课程中学习到的手工或自动迁移方式，将上述脚本迁移到昇腾 AI 处理器上，不要求执行训练，仅需提交迁移后的脚本。【分值：10 分】
3. 将上述迁移好的 LeNet 网络使用 mnist 数据集在 ModelArts 平台上正常跑通。【分值：10 分】
 - Pycharm 控制台界面有正常训练日志打印，给出截图。【子分值：5 分】
 - 将最终训练的模型权重文件（训练步数不限）保存在 OBS 上，给出截图。【子分值：2 分】
 - 给出当前训练的 CANN 运行日志截图，给出截图。【子分值：3 分】

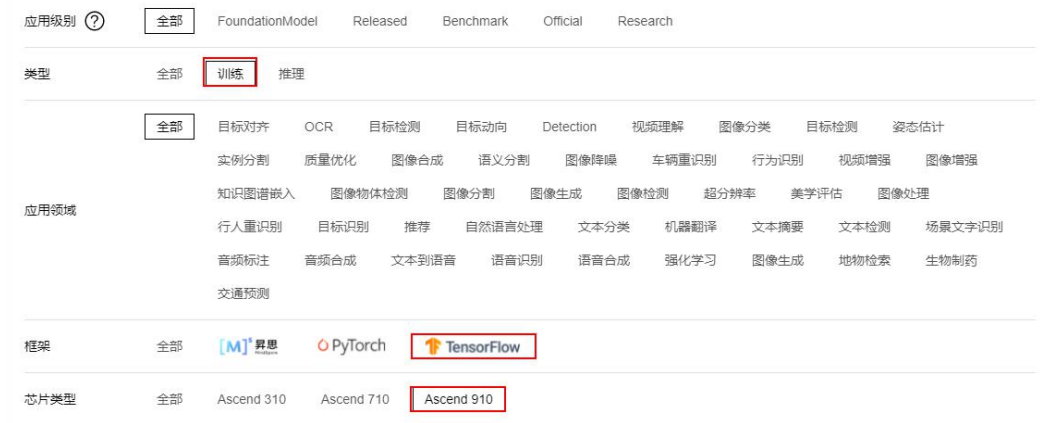
5.2 附加题

开发者自己选择昇腾社区 ModelZoo 的 Tensorflow 模型，迁移成功并在 ModelArts 平台上训练跑通。【分值：10 分】

- 迁移成功，提交迁移后的脚本。【子分值：5 分】
- 在 ModelArts 平台训练跑通，给出相关日志截图。【子分值：5 分】

友情提醒：

1. 昇腾社区 ModelZoo 的 Tensorflow 模型列表可单击 [LINK](#)，并通过如下筛选条件查找：



2. 没有迁移或训练成功也没关系，可在 [ModelZoo](#) 仓提 issue，也能酌情给分。

6

附录：作业提交格式

学员须在 <https://bbs.huaweicloud.com/forum/thread-181758-1-1.html> 下回帖，回帖格式：

新手班模型开发课小作业 1 提交 + 华为云账号
新手班模型开发课小作业 2 提交 + 华为云账号
新手班模型开发课小作业 3 提交 + 华为云账号
新手班模型开发课大作业提交 + 华为云账号
新手班模型开发课附加题提交 + 华为云账号

请按作业要求上传答案，如果涉及截图和源码等，请以压缩包的方式上传。