

MindSpore 训练模型推理精度报告

实现的模型

inceptionv4

提交结果

	Top1-Error	Top5-Error	Performance (单卡+batchsize=32)
baseline	17.7%	3.8%	355.00ms/step
我们的实现(完整 50000 张图片)	17.22%	3.79%	208.46ms/step (1.70 倍)
我们的实现(去掉 blacklist 剩下 48238 张图)	16.59%	3.45%	

注 1: 验证采用论文中用到的 144crop 的方法, 实现方式参考

<https://arxiv.org/pdf/1409.4842.pdf>, 所使用的 4 个 resize 大小为 320、352、384、448, 具体实现主要在 inceptionv4-adsl007-final/inceptionv4-adsl007-final-eval/model.py 中

注 2: blacklist 参考 https://github.com/zjytkz/ai-practice/blob/master/ILSVRC2015_devkit/ILSVRC2015/devkit/data/ILSVRC2015_clsloc_validation_blacklist.txt

网络训练速度

	单卡实验		8 卡实验	
	ms/step	samples/second	ms/step	samples/second
baseline	355	90	无	无
无优化	251(cb)	127(cb)	1128(cb)	907(cb)
开启数据下沉	208	153	940(cb)	1083(cb)
MindRecord 加速数据读取	待做	待做	待做	待做

注 1: 8 卡数据并行时每张卡 batchsize=128, 单卡训练时 batchsize=32

注 2: 每项实验重复 3 次, 数据去除 warmup 的 3 个 epoch, 取 10epoch 平均

注 3: 性能优化选项说明:

- 单卡无优化: 包含完整模型和数据预处理部分
- 开启数据下沉: 开启 dataset_sink_mode 选项
- 转化数据格式: 预先将 images 数据处理为 mindspore record 格式
- (cb)表示训练中开启 checkpoint_callback 的结果

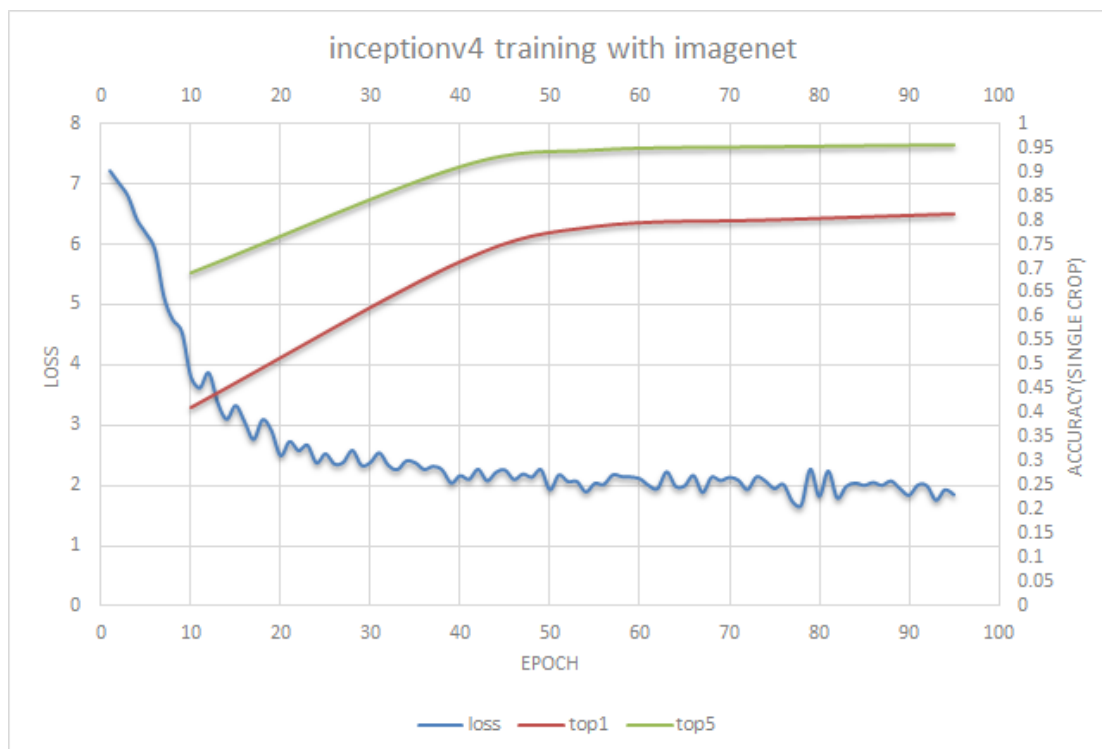
网络模型脚本、网络训练脚本、网络推理脚本

```
./
├── inceptionv4-adsl007-final
│   ├── inceptionv4-adsl007-final-train # 训练脚本目录
│   │   ├── inceptionv4 # 定义网络模型
│   │   ├── dataset_imagenet.py # 定义数据集
│   │   ├── train_inceptionv4_for_1_card.py # 单卡训练脚本
│   │   ├── train_inceptionv4_for_8_card.py # 八卡训练脚本
│   │   └── train_inceptionv4eval.py # single crop 验证脚本
│   └── inceptionv4-adsl007-final-eval # 144crop 验证脚本目录
│       ├── inceptionv4 # 定义网络模型, 同上
│       ├── dataset_imagenet.py # 定义数据集, 适应 144crop
│       ├── evaluation_144_crop.py # 144crop 验证脚本
│       └── model.py # 修改 mindspore.model 以支持 144crop
└── MindSpore 训练模型推理精度报告.pdf # 本文档
```

训练超参

1. 8 卡数据并行, 每张卡 batch_size = 128
2. epoch 1 - 40: 前 5 个 epoch 的 learning rate 从 0 增加到 0.2, 之后每 2 个 epoch, lr 乘以 0.94
3. epoch 41 - 75: 前 2 个 epoch 的 learning rate 从 0 增加到 0.05, 之后每个 epoch, lr 乘以 0.95
4. epoch 76 - 95: 前 2 个 epoch 的 learning rate 从 0 增加到 0.002, 之后每个 epoch, lr 乘以 0.95
5. loss function: SoftmaxCrossEntropyWithLogits, sparse=True, smooth_factor=0.1, reduction="mean", is_grad=False
6. optimizer: RMSProp, decay=0.9, epsilon=1.0

精度报告



可以看出：训练过程中的 single crop 输出的 evaluation 准确度，我们已经收敛得比 paper 中的结果更好了。论文中使用 single crop 在去掉 blacklist（剩下 48238 张图）的验证集上达到了 top1 80%， top5 95%的精度，而我们达到了 top1 81%， top5 95.5%。

checkpoint

obs://mindspore-res-commit-adsl/inceptionv4-adsl007-final-version.ckpt

obs://mindspore-res-commit-adsl/inceptionv4-adsl007-final-version.meta

如何在 ModelArts 上复现训练

将 inceptionv4-adsl007-final/inceptionv4-adsl007-final-train 文件夹下的所有文件上传至 obs://togo-obs/inceptionv4_imagenet/（根据自己需要修改对应自己的 obs 桶名称）下，在 ModelArts 上提交任务，填写如下表所示的任务参数：

任务参数	配置信息
代码目录	/togo-obs/inceptionv4_imagenet/
启动文件	/togo-obs/inceptionv4_imagenet/train_inceptionv4_for_8_card.py
数据存储位置	/mindspore-dataset/cifar-10 （必填参数、随便填一个）

训练输出位置	/mindspore-dataset/output/
运行参数	实际训练根据“训练超参”中的说明来修改源程序中的配置
作业日志路径	/mindspore-dataset/log/
资源池	公共资源池
规格	Ascend:8*Ascend 910 CPU:192 核 768GIB（或者单卡,将启动文件改为/togo-obs/inceptionv4_imagenet/train_inceptionv4_for_1_card.py）
计算节点个数	1

如何在 ModelArts 上复现推理

1. 144 crop 验证：将 inceptionv4-adsl007-final/inceptionv4-adsl007-final-eval 文件夹下的所有文件上传至 obs://togo-obs/inceptionv4_imagenet/（根据自己需要修改对应自己的 obs 目录）下，在 ModelArts 上提交任务，填写如下表所示的任务参数：

任务参数	配置信息
代码目录	/togo-obs/inceptionv4_imagenet/
启动文件	/togo-obs/inceptionv4_imagenet/evaluation_144_crop.py
数据存储位置	/mindspore-dataset/cifar-10（必填参数、随便填一个）
训练输出位置	/mindspore-dataset/output/
运行参数	无，在源文件中指定需要验证的 checkpoint 路径和文件
作业日志路径	/mindspore-dataset/log/
资源池	公共资源池
规格	Ascend:1*Ascend 910 CPU:192 核 768GIB
计算节点个数	1

注 1：完整 144crop 需要跑大约 80 小时

2. Single crop 验证：将 inceptionv4-adsl007-final/inceptionv4-adsl007-final-train 文件夹下的所有文件上传至 obs://togo-obs/inceptionv4_imagenet/（根据自己需要修改对应自己的 obs 目录）下，在 ModelArts 上提交任务，填写如下表所示的任务参数：

任务参数	配置信息
------	------

代码目录	/togo-obs/inceptionv4_imagenet/
启动文件	/togo-obs/inceptionv4_imagenet/train_inceptionv4eval.py
数据存储位置	/mindspore-dataset/cifar-10 （必填参数、随便填一个）
训练输出位置	/mindspore-dataset/output/
运行参数	无，在源文件中指定需要验证的 checkpoint 路径和文件
作业日志路径	/mindspore-dataset/log/
资源池	公共资源池
规格	Ascend:1*Ascend 910 CPU:192 核 768GIB
计算节点个数	1