

深度神经网络加速: cuDNN 与 TensorRT



班主任

悦雯





### ▶第一部分:上课注意事项

- ▲ 如何上课和交作业
- ▲ 课程安排
- ▲ 评优标准
- ▲ 学习守则
- ▲ 答疑规则

人工智能基础 , 机器学习 202301期

### 深度神经网络加速: cuDNN 与 TensorRT

- ◎ 内容精简: 主讲CUDA核心并行运算操作
- 知识前沿: 涵盖当下主流的深度学习模型加速工具
- ◎ 気围活跃:与数百位同学交流学习



🤦 讲师:杨伟光 👤 讲师:康博



▶第二部分:助教分享环节

## 如何上课



#### ●pc端 (推荐)

- ➤ step1:打开深蓝学院网站:https://www.shenlanxueyuan.com/,登录自己的深蓝账号,然后在右上角找到 "我的课程";
- > step2:进入之后在"我的课程"列表里, 找到本门课程,点击课程开始学习。



# 如何上课



#### ●手机APP (手机、平板电脑)

> 安装 "深蓝学院" APP, 可以在手机客户端观看视频或者下载课件。

注意:完成作业是需要在PC端完成的。



深蓝学院APP

# 如何交作业



- step1: 进入课程页面,在课程目录,每章节的最后一小节是作业;
- step2:点击【作业】任务,在线上传,提交作业。



7-1: 【作业】第一章

# 课程安排



第1章: CUDA C编程及GPU基本知识

第2章: CUDA C编程: 矩阵乘法

第3章: cuda stream 和 Event

第4章: cuDNN与cuBLAS

第5章: TensorRT介绍

第6章: TensorRT plugin用法

第7章: TensorRT量化加速

第8章: TensorRT 转换 ONNX模型

第9章:模型推理经验

课程内容: 见目录

学习方式:解锁制学习

课程有效期: 1年

服务周期: 6个月 (答疑、批改作业)

### 学习方式



#### ●课程视频解锁要求

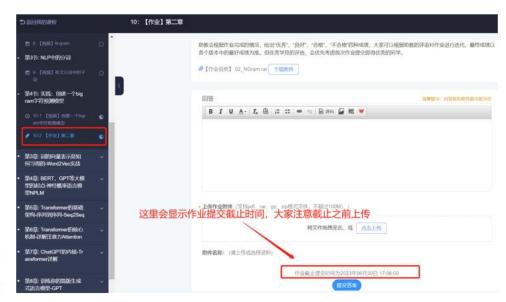
▶当前视频任务需要至少学习50%才能解锁一下小节的视频。

#### ●作业提交时间规则

▶系统会根据大家报名课程的时间自动设置截止时间。

#### ●作业提示解锁要求

▶各章节作业更新之后会定期上传思路提示的文档或视频,但需要大家在提交作业之后才能解锁对应章节的作业讲解。



### 作业要求



- **对于知识的真正理解在于动手实践。因此,学院课程作业中包含大量编程实践题目。**
- 学院不希望只培养对调库调参乐此不疲的同学。因此,实践作业中往往会包含不允许直接调库的要求, 而是要求同学们自己复现算法。
- 学院的视频时长与作业量基本按照1:6的比例去配置。因此,每周的课程作业基本需要花费1-2天的时间去完成。
- 为了培养大家的动手能力,所以作业均不提供源代码 (参考答案)。
- 作业批改频率为2天左右批阅一次,按提交顺序依次批改。

### 毕业和优秀



- 怎样才能毕业?
- ✓ 每章作业都提交并且达到 "合格"及以上,就可以拿到深蓝学院颁发的毕业证书。
- 获得优秀学员的条件是什么?
- ✓ 达到毕业条件的基础上,70%及以上的作业成绩为"优秀"可拿到优秀学员证书。
- 成为深蓝学院的优秀学员有什么好处?
- ✓ 你将获得一张带有你名字的优秀学员证书,放到简历上,HR看到会加分;
- ✓ 你报名其他课程时可以来找班主任领取50元优惠券;
- ✓ 优秀学员将有机会成为深蓝学院助教,跟讲师近距离沟通,得到学院潜在福利和更多职业发展机会。

# 答疑规则



#### ●讨论区

●主要用途:关于课程内容的答疑

复杂内容提问首选讨论区,便于回顾

#### ●微信群

●主要用途:即时性的讨论和交流

- ●禁忌
  - 调侃要有度,这里对任何形式的人身攻击,诱导引战以及不健康的内容持零容忍态度
  - 不得发与课程无关的信息,例如招聘信息、广告或外链, 若未征求班主任许可,擅自在群里发以上信息,直接移出群聊
  - 社群管理员保持清理一切不合规信息行为的权利



#### 讨论区使用方法:

观看视频过程中遇到疑问可以点击【问答】 功能提问,问题会在讨论区展示,为了方便 老师快速定位问题所在作出解答,提交时记 得勾选时间戳。

(ps:复杂问题优先选择讨论区,其他即时性

问题可以选择答疑群)







### 感谢各位聆听 Thanks for Listening

