

2024-2025 秋季学期“深度学习和计算机视觉”课程实验三：

CIFAR-ViT

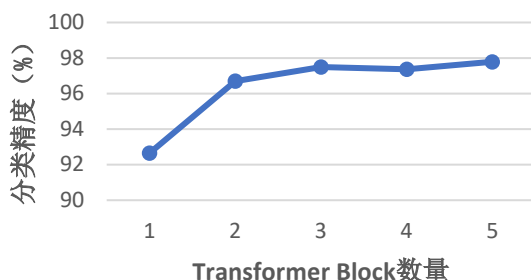
截止日期：2024 年 10 月 31 日 23: 59

1. 实验背景

- **课程内容回顾** 我们重点学习了自注意力机制与 Transformer，介绍了 Vision Transformer (ViT) 架构在图像分类任务的应用
- **实验内容** 利用 PyTorch 等框架，设计和搭建 ViT 实现 CIFAR-10 数据集分类任务，分析 ViT 中各模块作用与工作机理；探索降低 ViT 复杂度的方法。（**数据集在群文件自行下载**）

2. 实验要求

- 2.1（必做）搜索和阅读相关资料和论文，实现 ViT 的构建，在**本课程提供的 CIFAR-10 数据集**完成分类任务（**在 CIFAR10_imbalanced 训练，在 CIFAR10_balance 测试**）
- 2.2（必做）训练数据集存在**类不平衡现象**，尝试至少 2 种解决方法并做出解释。
- 2.3（必做）分析 ViT 不同模块，如 patch size、embedding dimension、多头注意力机制数量、Transformer block 数量，对分类结果的影响（参考绘制并对比不同参数设定下的收敛曲线）



- 2.4（选做）在确保分类性能不明显下降的前提下，尝试 ViT 轻量化（如 Dynamic ViT、A ViT 等，不少于 2 种方法），**提升训练/推理速度，降低模型参数量。改变 2.3 中的模块参数不算轻量化方法。**
- 提交实验报告，简要叙述 ViT 工作流程，实验步骤，实验结果分析；需明确说明组员分工、给出组内贡献（共计 100%，如甲 40%，乙 30%，丙 30%）
- 评分标准：必做项目占 80 分，选做项目占 20 分
- 提交代码¹

¹复杂度计算可采用 torchsummary 或其他库，计算模型参数量和 FLOPs