

课后作业

手写或打印一页A4纸/题

邮件主题 HW-9-姓名

作业编号 9

满分 6+6 分

Q1 (6分) ViT 代码理解

下载 [code](#)，建议上传至 google colab. 使用 GPU 运行。

1. 阅读readme及vit.ipynb代码。
2. 下载数据（本地建议手动下载，Colab建议使用download=True）。
3. 理解einops对图像快速chunk，也就是tokenize。
4. 理解QKV键值对attention的实现方式，理解multi-head的实现方式。
5. 理解相关矩阵correlation并可视化。
6. 回答所有TODO-Explain内容。

Q2 (6分) RoPE 位置编码

修改 Net 网络定义中，`get_sinusoidal_positional_encoding()` 替换改为 RoPE-Attention。

1. 阅读RoPE的[文章](#)。同时，拓展到ViT的[RoPE-2D](#)文章，理解(7)-(15)，参考其实例[代码](#)。
2. 回答为什么Rope是在每一层QK上进行作用，而不是在输入x上作用。
3. 修改实现RoPEAttention代码，并重新训练ViT-RoPE，查看accuracy以及attention。哪一种位置编码更好？