广东工业大学

先进制造学院

期末大作业



|  |  |
| --- | --- |
| 课程名称： | 深度学习 |
| 姓名学号： |  |
| 专业班级： |  |
| 填写日期： |  |

**基于Transformer的Image Caption Generation**

*（题目和模版都仅供参考，可根据需要自由发挥）*

**一、引言**

**1.1 研究背景与意义**

阐述 Image Caption Generation 任务对计算机视觉和自然语言处理的重要性，说明 Transformer 解决该任务的优势、研究意义。

**1.2 研究目标与内容**

明确使用 Transformer 解决 Image Caption Generation 任务的具体目标和内容。

**二、相关技术原理**

**2.1 Image Caption Generation 任务概述**

解释任务定义，分析其在实际应用中的重要性和挑战。

**2.2 Transformer 模型原理**

**2.2.1 架构组成**

介绍 Transformer 的编码器、解码器结构，以及各模块的作用。

**2.2.2 自注意力机制**

阐述自注意力机制的原理和计算过程，说明其在捕捉长距离依赖关系方面的优势。

**2.2.3 多头注意力机制**

讲解多头注意力机制如何并行处理不同表示子空间，提升模型对复杂关系的理解能力。

**2.3 注意力权重可视化原理**

介绍将注意力权重以热力图等形式展示的原理和作用，说明其对理解模型决策过程的帮助。

**三、模型构建**

**3.1 数据描述**

对 COCO进行简要介绍。

**3.2 图像特征提取**

介绍使用 VGG预训练模型提取图像特征的方法

说明删除最后若干层全连接层的原因

**3.3模型搭建**

首先画出完整的模型结构图。

**3.3.1 编码器设计**

说明编码器如何将图像特征转换为隐藏表示，分析其对图像信息的编码能力。

**3.3.2 解码器设计**

介绍解码器如何根据隐藏表示逐步生成字幕，说明生成过程中的关键步骤。

**3.3.3 模型训练与优化**

阐述所采用的损失函数和优化算法进行训练的过程，分析模型的收敛情况。

**四、实验与结果分析**

**4.1 实验设置**

介绍实验使用的硬件和软件环境。

**4.2 文生图实验结果分析**

展示模型在测试集上生成的字幕示例，以及各项评估指标的数值结果。分析实验结果

**4.3 注意力权重提取及可视化结果分析**

说明在模型训练或推理过程中，如何从自注意力层提取注意力权重。

分析可视化结果中注意力在图像区域和文本生成过程中的分布规律。

探讨注意力分布与模型生成字幕准确性和合理性之间的关系。

**五、结论与展望**