

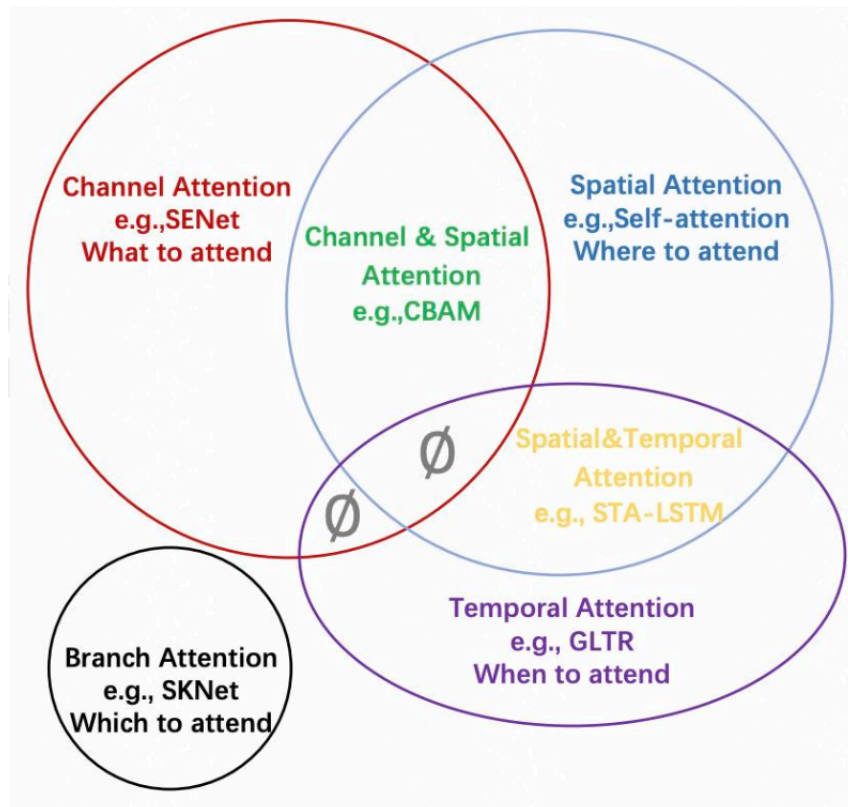
# 《深度学习平台与应用》作业三

## 20251130

### 一、选择题

1. 关于循环神经网络（RNN）及 LSTM（Long Short Term Memory），下列说法中不正确的是：
  - A. 在标准的 LSTM 架构中，遗忘门（Forget Gate） $f_t$  决定了上一时刻的状态  $C_{t-1}$  有多少信息被保留到当前时刻。
  - B. 相比于 Vanilla RNN，LSTM 通过引入门控机制和细胞状态（Cell State） $C_t$  的加法更新，能够缓解梯度消失问题，从而更容易捕捉长距离依赖。
  - C. 在 LSTM 的更新公式中，输出门（Output Gate） $o_t$  控制了当前隐藏状态  $h_t$  的输出，其计算公式通常涉及 Sigmoid 激活函数。
  - D. LSTM 中计算当前时刻隐藏状态  $h_t$  的最终公式为  $h_t = o_t \odot \sigma(C_t)$ ，其中  $\sigma$  代表 Sigmoid 函数。
2. 关于 Transformer 架构及其核心组件“注意力机制”（Attention Mechanism），下列说法正确的是：
  - A. 在自注意力（Self-Attention）计算中，除以缩放因子  $\sqrt{d_k}$  的主要目的是为了放大点积结果，从而在反向传播时获得更大的梯度。
  - B. Transformer 模型完全抛弃了递归结构（Recurrence），为了让模型利用序列的顺序信息，必须在输入嵌入（Input Embeddings）中加入位置编码（Positional Encodings）。
  - C. 在 Transformer 的 Decoder 中，Masked Multi-Head Attention 的作用是让模型在训练时能够看到当前位置之后的所有单词，从而更好地进行上下文建模。
  - D. 视觉 Transformer（ViT）将图像分割成多个 Patch，直接将这些 Patch 展平后输入到 Transformer Encoder 中，不需要任何额外的线性投影（Linear Projection）或位置编码。

3. 根据对注意力机制分类的总结，以下关于不同类型注意力机制的描述，正确的是：



- A. SENet (Squeeze-and-Excitation Networks) 是一种典型的空间注意力机制 (Spatial Attention)，主要关注“Where to attend”。
- B. CBAM (Convolutional Block Attention Module) 仅包含通道注意力，不包含空间注意力。
- C. Self-attention (自注意力机制) 属于空间注意力 (Spatial Attention) 的一种，主要关注特征图的 H (高度) 和 W (宽度) 维度上的依赖关系。
- D. 所有的注意力机制都必须同时包含通道 (Channel)、空间 (Spatial) 和时间 (Temporal) 三个维度的建模。
4. 在 Transformer 架构中，Positional Encoding (位置编码) 的主要作用和实现方式是：
- A. 通过卷积层提取位置特征，替换原有的 Input Embedding。
- B. 将位置信息与 Input Embedding 进行拼接 (Concatenate)，以增加特征维度。
- C. 因为 Self-Attention 机制本身无法捕捉序列的顺序信息，所以需要将位置编码与 Input Embedding 相加 (Add)，且通常使用正弦

- 和余弦函数生成。
- D. 位置编码是一个可学习的权重矩阵，必须通过反向传播从头训练，不能使用固定的数学公式。
5. 关于 Vision Transformer (ViT) 如何将二维图像处理为 Transformer 能够接受的输入格式，下列描述正确的是：
- A. 使用预训练的 CNN（如 ResNet）提取特征图，直接作为 Transformer 的输入。
  - B. 将整张图片拉伸为一个长向量，直接进行线性投影。
  - C. 将图像分割成固定大小的 Patches（例如  $16 \times 16$ ），将每个 Patch 展平后进行线性投影（Linear Projection），并加上位置编码。
  - D. ViT 不需要位置编码，因为图像块本身具有空间结构。
6. 多头注意力（Multi-Head Attention）相比单头注意力的核心优势是什么？
- A. 减少参数量，提升训练速度
  - B. 允许模型并行关注不同子空间中的关系
  - C. 避免使用缩放点积注意力
  - D. 使注意力矩阵稀疏化，从而减少计算量
7. 在目标检测与分割模型的发展历程中（R-CNN 系列），针对特征对齐和计算效率的改进是关键。以下关于这些模型描述，错误的是：
- A. Fast R-CNN 引入了 RoI Pooling 层，使得我们不再需要对每个 Region Proposal 分别进行卷积计算，而是共享整张图的特征图，大大提高了训练和推理速度。
  - B. Faster R-CNN 的核心创新在于提出了区域生成网络（RPN），它是一个全卷积网络，可以与检测网络共享卷积特征，实现了端到端的训练。
  - C. Mask R-CNN 在 Faster R-CNN 的基础上增加了一个预测分割掩码（Mask）的分支，用于实例分割任务
  - D. 为了解决特征图与原始图像区域对齐的问题，Mask R-CNN 沿

用了 Fast R-CNN 中的 RoI Pooling 操作，以保证特征提取的精确性。

8. R-CNN 被认为“非常慢”的主要原因是什么？
- A. CNN 只能处理灰度图
  - B. SVM 分类器训练困难
  - C. 对每个 proposal 都独立运行 卷积网络处理，重复大量计算
  - D. 特征图分辨率太低，导致需要重复补偿计算

## 二、问答题

1. 关于循环神经网络：

(a): 请描述循环神经网络（RNN）的主要结构和流程。

(b): RNN 展开之后的计算图长什么样？为什么说“各时间步共享参数”？

(c): 请分别举例说明 RNN 可以处理的几种输入/输出结构，以及对应的常见任务。

(d): 为什么RNN速度慢。

(e): 为什么RNN有梯度问题。

(f): LSTM 相比普通 RNN，在结构上增加了哪些“门”？它们各自控制什么？

(g): 为什么说 LSTM 更有利于缓解RNN 的梯度问题？

## 2. 关于Transformer模型:

(a): Transformer中有哪些关键组成。

(b): 请解释为什么Transformer中需要位置编码，位置编码是怎么添加的。

(c): 请解释self-attention机制的流程。

(d): self-attention 和 multi-head attention 有哪些区别，multi-head attention 有什么好处。

(e): Feed forward 层在Transformer中的作用是什么。

(f): Add & Norm 在Transformer中的作用是什么。

(g): 比较卷积网络，双向循环网络，全连接网络和注意力机制的异同。

(h): 什么是mask attention以及mask attention在Transformer中是如何应用的。

3. 关于目标检测和图像分割:

(a): 语义分割与目标检测的主要区别是什么。

(b): 什么是全卷积网络? 它如何解决语义分割中的高分辨率问题?

(c): 目标检测中如何解决“多个物体”的问题?

(d): 什么是R-CNN? 它如何提升目标检测的效率?

(e): Fast R-CNN 与 R-CNN 的区别是什么? 为什么Fast R-CNN更高效?

(f): 实例分割和语义分割有什么区别。

(g): Mask R-CNN 在目标检测的基础上, 如何扩展为实例分割任务?