



Transformer 代码从零解读

【代码解析】



扫码关注微信公众号

文章周更

知识分享

一起进步

求关注，求点赞，求一切！！

三类应用

1. 机器翻译类应用-Encoder和Decoder共同使用
2. 只使用Encoder端-文本分类BERT和图片分类VIT
3. 只使用Decoder端-生成类模型

整体看是这样的

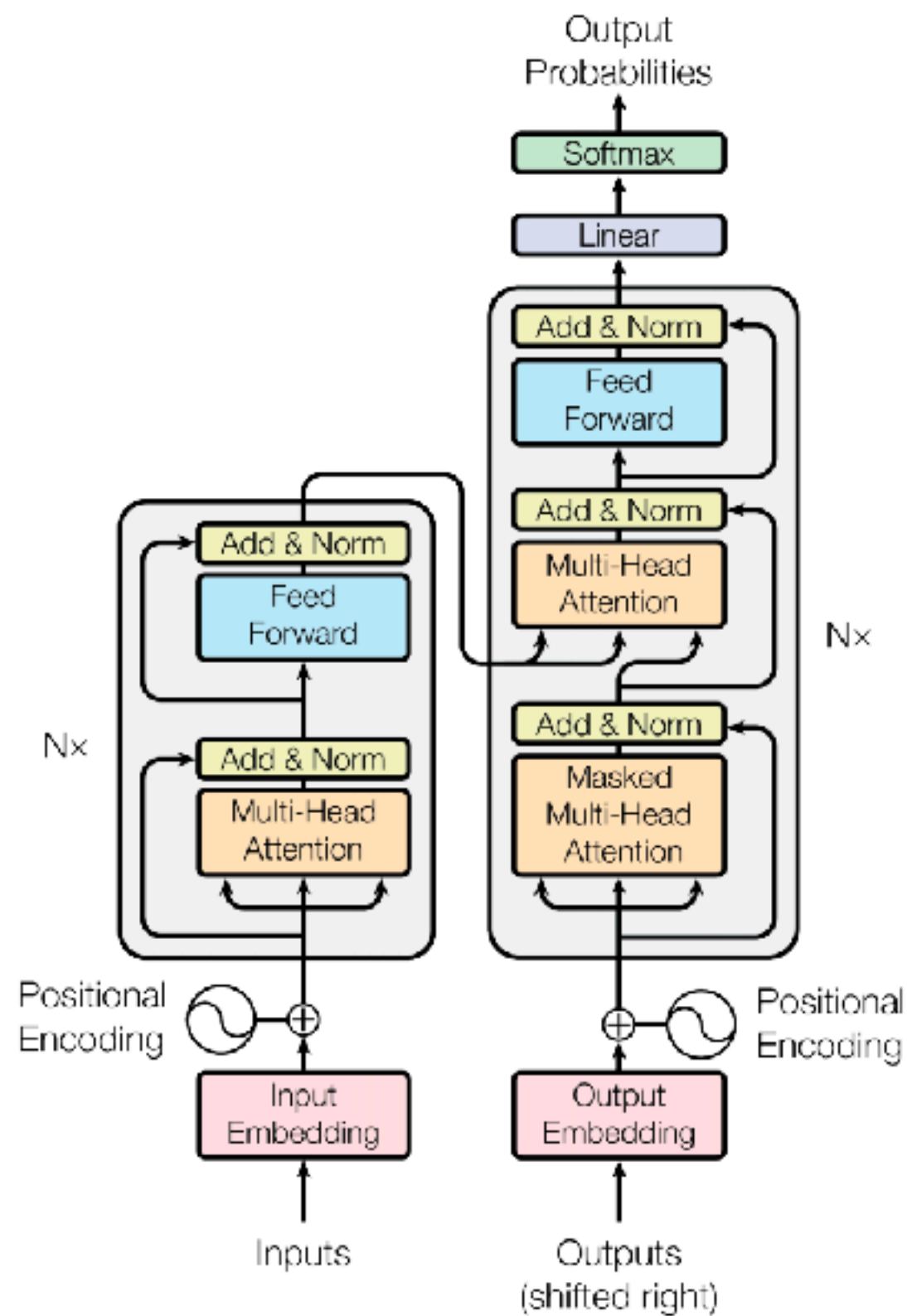
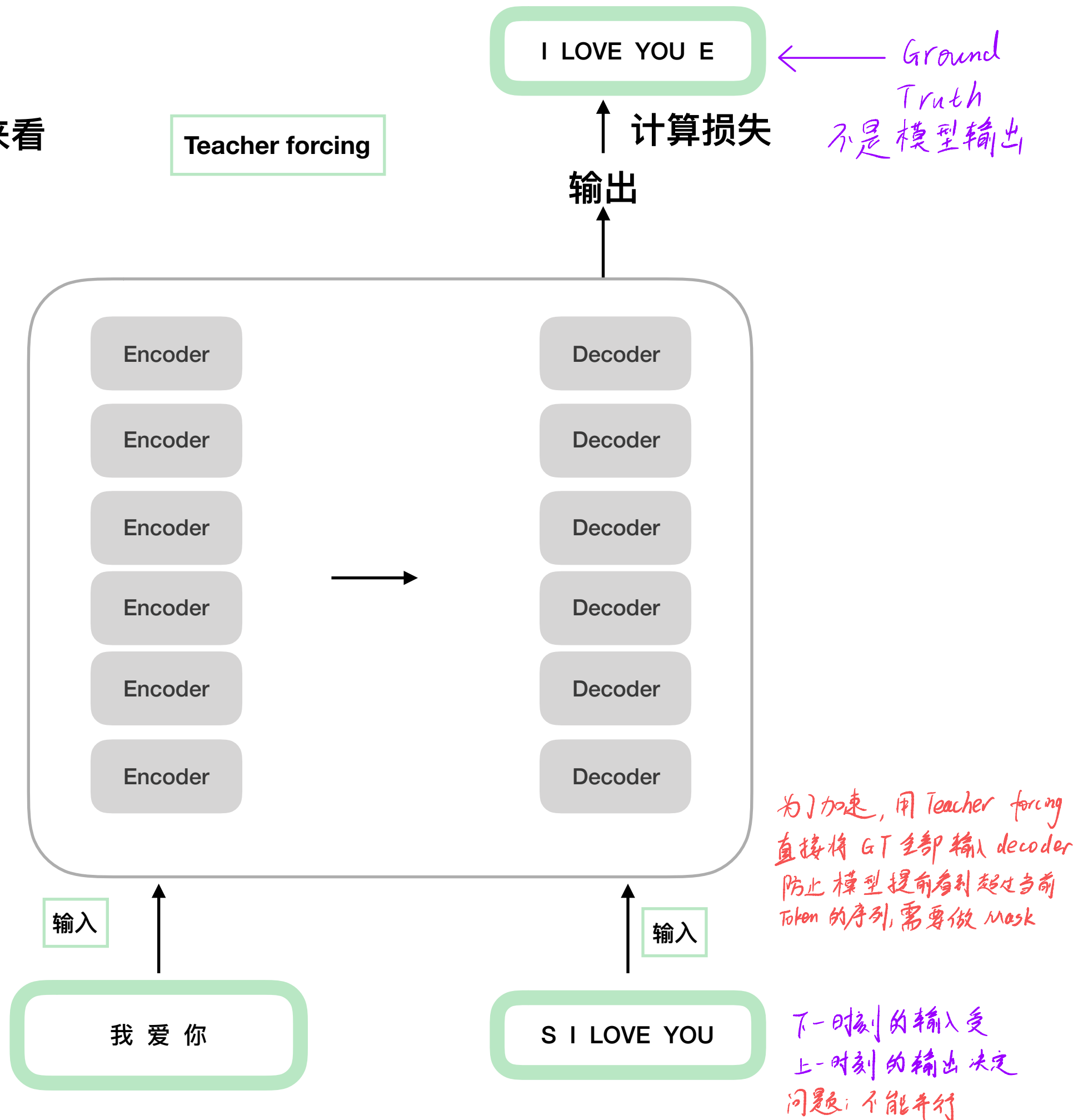
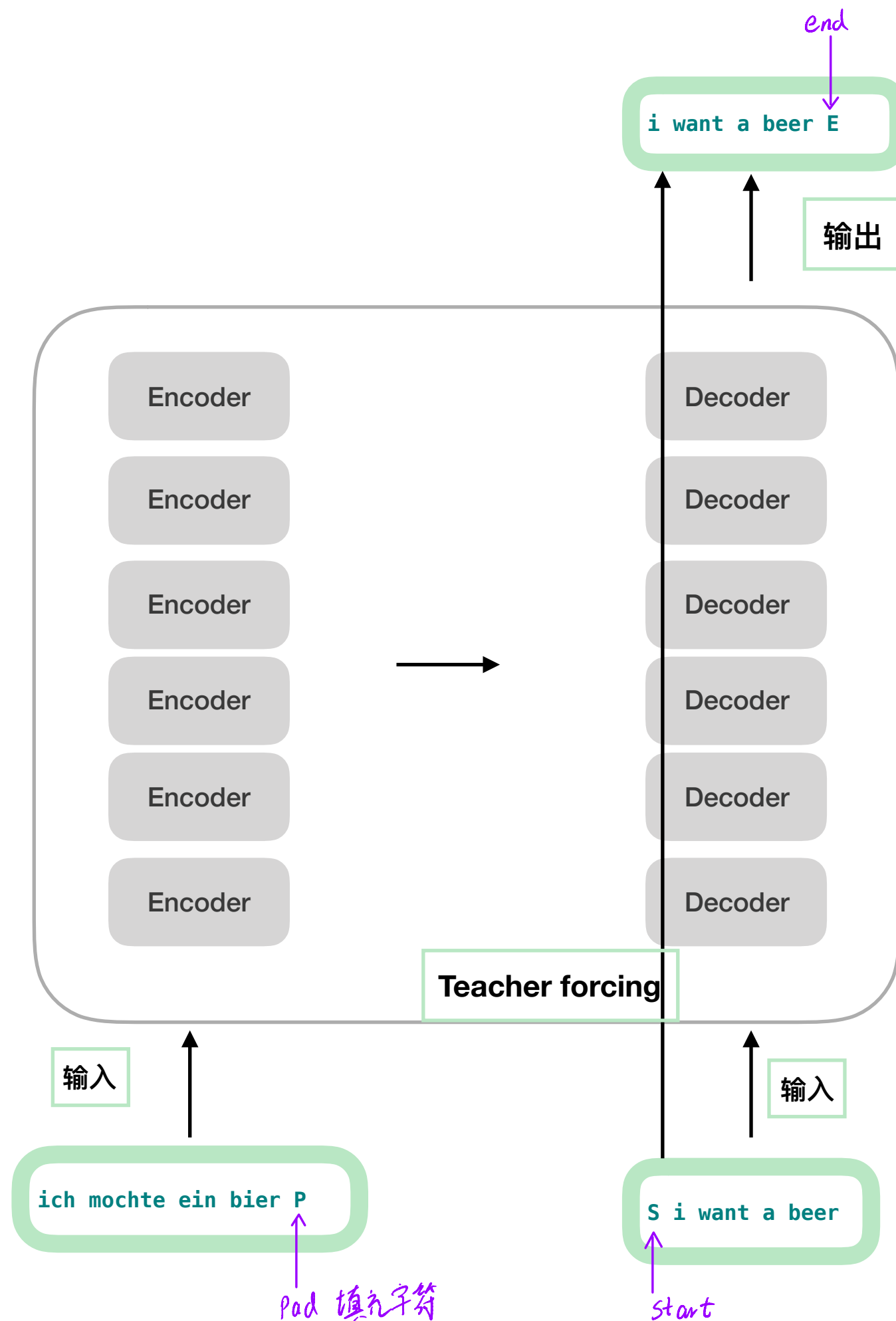


Figure 1: The Transformer - model architecture.

我们抽离出来看





句子真实长度为

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 别 | 休 | 息 | , | 卷 | 起 | 来 | | |
| 今 | 天 | 天 | 气 | 真 | 不 | 错 | 啊 | |
| 大 | 家 | 好 | , | 都 | 吃 | 饭 | 了 | 吗 |
| 真 | 不 | 错 | 哈 | | | | | |

7

8

9

4

假设max_len=8

组成有效矩阵

句子真实长度为

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 别 | 休 | 息 | , | 卷 | 起 | 来 | P | |
| 今 | 天 | 天 | 气 | 真 | 不 | 错 | 啊 | |
| 大 | 家 | 好 | , | 都 | 吃 | 饭 | 了 | 吗 |
| 真 | 不 | 错 | 哈 | P | P | P | P | |

7

8

9

4

复现代码心得体会

1. 从整体到局部

2. 搞清楚数据流动形状，非常关键

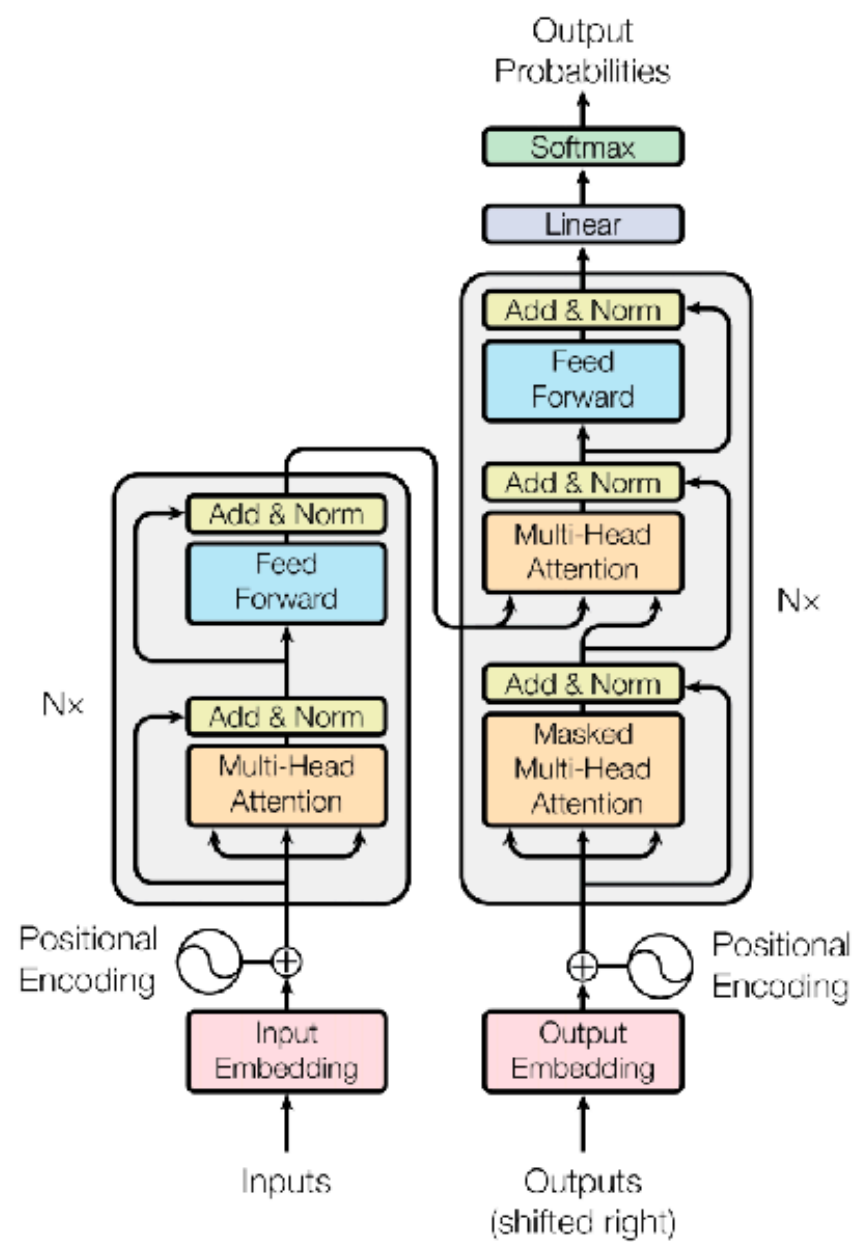


Figure 1: The Transformer - model architecture.

位置编码公式

512个维度

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{\text{model}}})$$
$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}})$$

两个共有的部分： $e^{-(2i)/d_{\text{model}} \cdot \log(10000)}$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

为什么需要告诉后面模型哪些位置被PAD填充

| | 卷 | 起 | 来 | PAD | |
|-----|----|----|----|-----|---------|
| 卷 | 20 | 5 | 4 | 9 | softmax |
| 起 | 5 | 30 | 8 | 12 | softmax |
| 来 | 4 | 8 | 15 | 14 | softmax |
| PAD | 9 | 12 | 14 | 40 | softmax |

| | 卷 | 起 | 来 | PAD |
|-----|----|----|----|-----|
| 卷 | 20 | 5 | 4 | 9 |
| 起 | 5 | 30 | 8 | 12 |
| 来 | 4 | 8 | 15 | 14 |
| PAD | 9 | 12 | 14 | 40 |



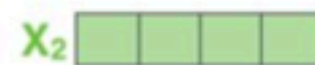
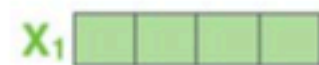
| 符号矩阵 | | | |
|------|---|---|---|
| 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 |

输入

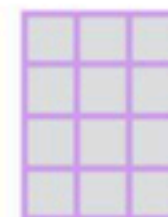
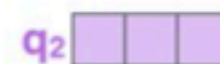
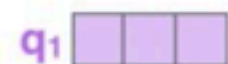
Thinking

Machines

词嵌入

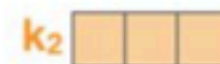
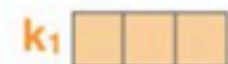


查询向量



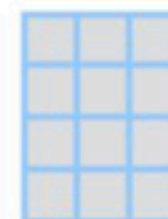
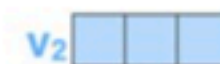
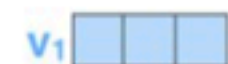
W^Q

键向量



W^K

值向量



W^V

自注意力的mask

| | 卷 S | 起 卷 | 来 起 | E 来 |
|---|--------|--------|--------|--------|
| S | 0 | 1 | 1 | 1 |
| 卷 | 0 | 0 | 1 | 1 |
| 起 | 0 | 0 | 0 | 1 |
| 来 | 0 | 0 | 0 | 0 |

decoder 其实有两个掩码
① PAD
② MASK