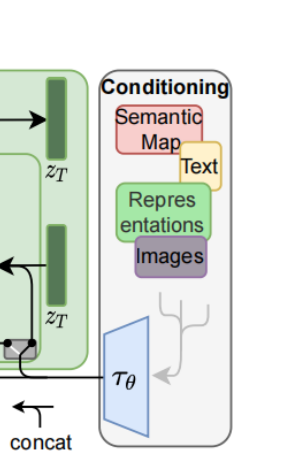
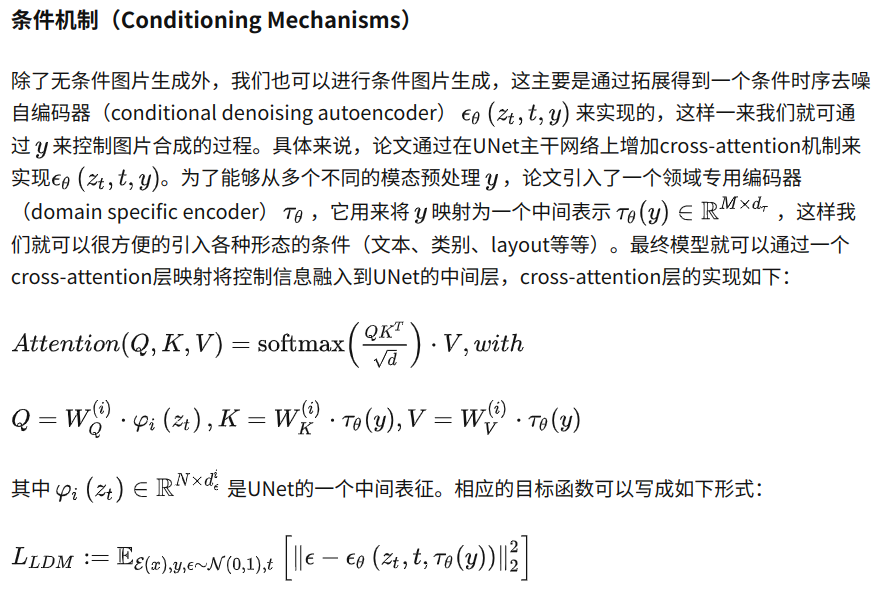
个人理解：



要回答的问题应该是这一部分，然后和左边的cross-attention机制。

下面这篇文章仔细看下，我感觉他说基本全的，有的部分可以深挖下。

<https://zhuanlan.zhihu.com/p/583124756>

对这篇文章的条件机制进行更详细的分析，比如这个领域专用编码器是什么，cross-attention层是什么，啥啥啥。这部分应该就包括了上面的两个问题了。