2024年0603-0607周报（张倩倩）

周总结

近期工作复盘（周五）

1.

2.

本周工作（周一制定）

1. CCF顶会论文调研。大批量，idea的。顺手整理成笔记，笔记是为了写论文方便来做。按照结构，引言和前面需要参考文献，实验部分需要代码参考。

2.计算资源跑起来。【周二周三，论文找完就该试试了】

3. PRCV会议的小目标竞赛。【平台跑起来以后，再弄弄】

4.周一先3篇，往比赛的那个方向进行调研以及大小模型，广泛阅读。【小模型3篇，大模型暂未】

5. 调研一下今年还能投递的顶刊的截止日期和相关文章。【暂时找到1个】

6. 可以认识一些新朋友。【收获大三涂涂】

下一步计划（周五）

1.

2.

3.

4.

5.

6.

面临的问题&生活杂事

6.11党支部中国考古博物馆带队中关村

6.24-6.25所里国际会议

6.28下午护照办理

0603

1.ACM MM会议

2024已经没戏了。

会议时间/地点：2024年10月28日至11月1日，澳大利亚，墨尔本  
截稿时间：2024年6月17日  
录用通知时间：2024年7月15日  
会议官网：<https://2024.acmmm.org/>

该会议发表的相关论文;

[1] Murari Mandal, Lav Kush Kumar, and Santosh Kumar Vipparthi. 2020. MOR-UAV: A Benchmark Dataset and Baselines for Moving Object Recognition in UAV Videos. In Proceedings of the 28th ACM International Conference on Multimedia (MM '20). Association for Computing Machinery, New York, NY, USA, 2626–2635. <https://doi.org/10.1145/3394171.3413934> （现有的 UAV 数据集主要侧重于物体检测。现有**物体检测器不会区分移动物体和非移动物体**。给定实时 UAV 视频流，我们如何定位和分类**移动**物体，即执行移动物体识别 (MOR)？......据我们所知，没有可用于 UAV 视频中 MOR 评估的标记数据集。因此，在本文中，我们介绍了 MOR-UAV，这是一个用于航拍视频 MOR 的大型视频数据集）

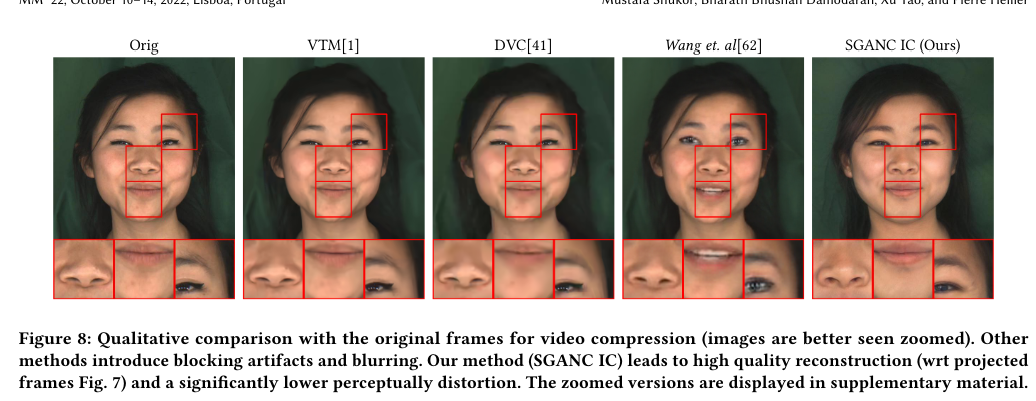
[2] Cai YuanQiang, Dawei Du, Libo Zhang, Longyin Wen, Weiqiang Wang, Yanjun Wu, and Siwei Lyu. 2020. Guided Attention Network for Object Detection and Counting on Drones. In Proceedings of the 28th ACM International Conference on Multimedia (MM '20). Association for Computing Machinery, New York, NY, USA, 709–717. <https://doi.org/10.1145/3394171.3413816> (提出一种新的引导注意力网络（GAnet），用于基于特征金字塔处理物体检测和计数任务。与以前依赖**无监督**注意力模块的方法不同，我们利用提出的背景和物体之间的**弱监督**背景注意力（BA）融合不同尺度的特征图，以获得更具语义的特征表示。然后，开发了前景注意力（FA）模块来考虑物体的全局和局部外观，以促进准确定位。此外，新的数据论证策略旨在在具有各种照明条件的无人机场景中训练出一个稳健的模型。在三个具有挑战性的基准（即 UAVDT、CARPK 和 PUCPR+）上进行的大量实验表明，与现有方法相比，所提出的方法具有最先进的检测和计数性能。代码可以在 https://isrc.iscas.ac.cn/gitlab/research/ganet 找到){ 哈哈作者王伟强是给我上过课的老师，一个很有趣的老师。偶然发现这篇文章。}【照他这样我也能写】

[3] Wei Gao, Hang Yuan, Yang Guo, Lvfang Tao, Zhanyuan Cai, and Ge Li. 2022. OpenHardwareVC: An Open Source Library for 8K UHD Video Coding Hardware Implementation. In Proceedings of the 30th ACM International Conference on Multimedia (MM '22). Association for Computing Machinery, New York, NY, USA, 7339–7342. https://doi.org/10.1145/3503161.3548543 ( 单位：北大深圳、鹏城实验室深圳。**8K超高清视频的硬件加速实时压缩**是最新视频编码标准赋能的典型应用。然而，新近发布的第三代音视频编码标准（AVS3）新增的编码工具大大增加了编码复杂度，严重阻碍了硬件编码器的高效实现。为了突破已知的瓶颈，本文提出了第一个用于8K超高清视频编码硬件实现的开源软件库OpenHardwareVC。具体来说，该库在分析原有**AVS3【央视春晚、国产的】**软件算法的基础上，给出了编码单元（CU）划分、帧内预测、变换和熵编码四个主要编码阶段的硬件加速设计。在Xilinx VU440 FPGA上的仿真结果表明，基于该库中封装的软件描述模块可以轻松支持30帧/秒（fps）的8K超高清视频实时压缩。该库的发布对超高清视频编码的硬件设计和系统实现非常有利，也有利于新编码标准的推广。OpenHardwareVC的开源库可以在https://git.openi.org.cn/OpenHardwareVC上找到。){果然只要涉及到AVS的论文都写的很短，要不就是写的和宣传广告一样}

以上是2020年的4年前的重点

2022年已经很少看到这些了，转向到了视频分类、视频压缩、多模态、跨模态还有剪枝和蒸馏。

[4] Mustafa Shukor, Bharath Bhushan Damodaran, Xu Yao, and Pierre Hellier. 2022. Video Coding using Learned Latent GAN Compression. In Proceedings of the 30th ACM International Conference on Multimedia (MM '22). Association for Computing Machinery, New York, NY, USA, 2239–2248. <https://doi.org/10.1145/3503161.3548219>(面部视频压缩。实验结果呈现方式启发了我)



[5] Yiming Sun, Bing Cao, Pengfei Zhu, and Qinghua Hu. 2022. DetFusion: A Detection-driven Infrared and Visible Image Fusion Network. In Proceedings of the 30th ACM International Conference on Multimedia (MM '22). Association for Computing Machinery, New York, NY, USA, 4003–4011. https://doi.org/10.1145/3503161.3547902 （红外与可见光图像融合旨在利用两种模态之间的互补信息来合成包含更丰富信息的新图像。现有的大多数工作都集中在如何更好地融合两种模态在对比度和纹理方面的像素级细节，而忽略了图像融合的意义在于更好地服务于下游任务。对于物体检测任务，图像中的物体相关信息往往比仅仅关注图像的像素级细节更有价值。为了填补这一空白，我们提出了一个检测驱动的红外和可见光图像融合网络，称为 DetFusion，它利用在物体检测网络中学习到的物体相关信息来指导多模态图像融合。我们将图像融合网络与两种模态的检测网络级联，并利用融合图像的检测损失为图像融合网络的优化提供任务相关信息的指导。考虑到物体位置为图像融合提供了先验信息，我们提出了一种物体感知的内容损失，以激励融合模型更好地学习红外和可见光图像中的像素级信息。此外，我们设计了一个共享注意力模块来激励融合网络从对象检测网络中学习特定于对象的信息。大量实验表明，我们的 DetFusion 在保持像素强度分布和保留纹理细节方面优于最先进的方法。更值得注意的是，在任务驱动评估中与最先进的图像融合方法的性能比较也证明了所提出方法的优越性。我们的代码将可用：https://github.com/SunYM2020/DetFusion。）**{对口了，行，我也能整}**

[6] Bin Yang, Mang Ye, Jun Chen, and Zesen Wu. 2022. Augmented Dual-Contrastive Aggregation Learning for Unsupervised Visible-Infrared Person Re-Identification. In Proceedings of the 30th ACM International Conference on Multimedia (MM '22). Association for Computing Machinery, New York, NY, USA, 2843–2851. [https://doi.org/10.1145/3503161.3548198](https://doi.org/10.1145/3503161.3548198()) （可见红外行人重新识别 (VI-ReID) 旨在从其他光谱相机捕获的图库集中搜索出相应的红外（可见）图像。最近的研究主要集中于有监督的 VI-ReID 方法，该方法需要大量跨模态（可见红外）身份标签，而这些标签比单模态行人重新识别中的注释更昂贵。对于无监督学习可见红外重新识别 (USL-VI-ReID)，较大的跨模态差异导致难以生成可靠的跨模态标签和学习没有任何注释的模态不变特征。为了解决这个问题，我们提出了一种新颖的增强双对比聚合 (ADCA) 学习框架。具体而言，提出了一种具有两个模态特定记忆的双路径对比学习框架来学习模态内行人表征。为了关联正向跨模态身份，我们设计了一个具有计数优先级的跨模态记忆聚合模块，以选择高度相关的正向样本，并在聚类级别聚合它们对应的记忆特征，确保优化明确集中在与模态无关的视角上。大量实验表明，我们提出的 ADCA 在各种设置下都明显优于现有的无监督方法，甚至超越了一些监督方法，从而促进 VI-ReID 在现实世界中的部署。代码可在 https://github.com/yangbincv/ADCA 获得.）

【但是上面这些都没有结合大模型】

2.ICCV2025会议（还能投递）

发稿方向：多模态学习+检测

会议时间/地点：2025年12月14日至20日，中国，深圳

截稿时间：待定

录用通知时间：待定

会议官网：https://iccv2025shenzhen.gitee.io/

遗留问题

1. 除了ICCV应该还有一些会议可以投递。边找边看。

0604

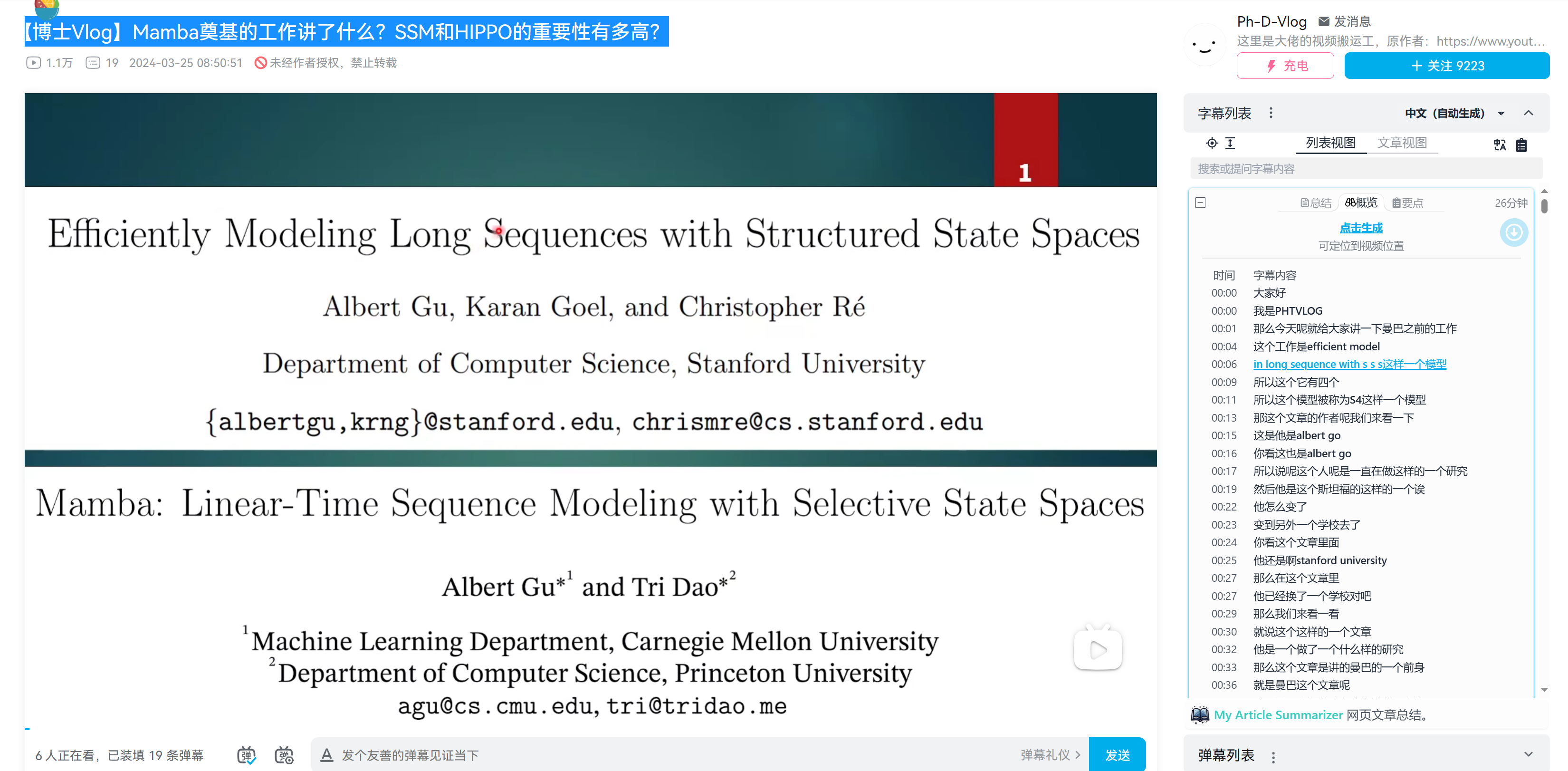
1.什么是Mamba模型【上午 那个word 碎片时间的那个】

基于状态空间模型（State Space Model）的Mamba模型最主要的***优势***在于其在**长序列任务**上的优异性能与较低的计算复杂度。【论文题目：Mamba: Linear-Time Sequence Modeling with Selective State Spaces】

（尝试看了论文、一些博客和微信公众号以及知乎。但是看完依旧不是很懂。有很多小概念不懂，有很多公式。所以，打开了哔哩哔哩。

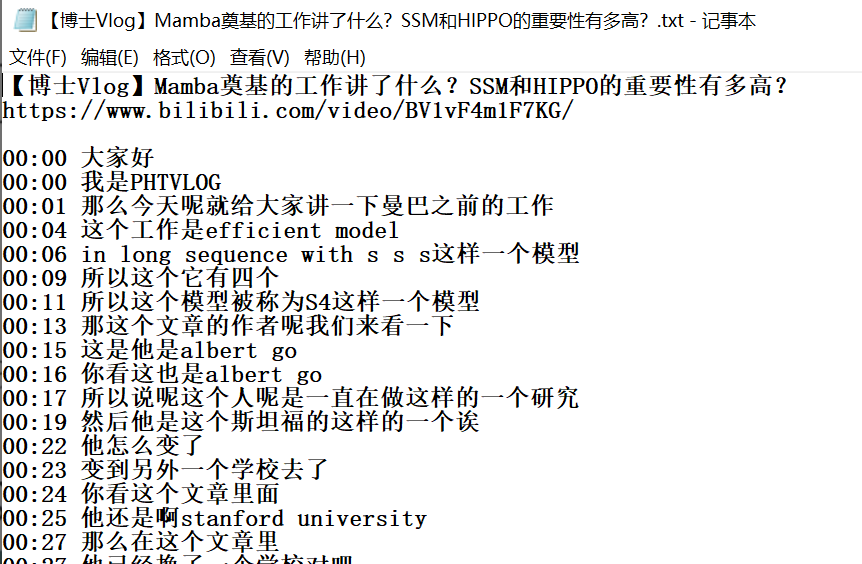
<https://search.bilibili.com/all?keyword=Mamba&from_source=webtop_search&spm_id_from=333.1007&search_source=5> ）

首先第一个视频：



（笔记见后）

【相应视频的字幕，我也添加到本周压缩包的txt里面了，相当于笔记，看文字能节省不少时间】



第二个视频：

https://www.bilibili.com/video/BV1KH4y1W7cm/?spm\_id\_from=333.337.search-card.all.click&vd\_source=1412c55689d7b78016c608ffdba64e83



（笔记见后）

。

2. 多模态小目标检测大模型相关论文【上午 那个word 碎片时间的那个 看有没有能跑的】

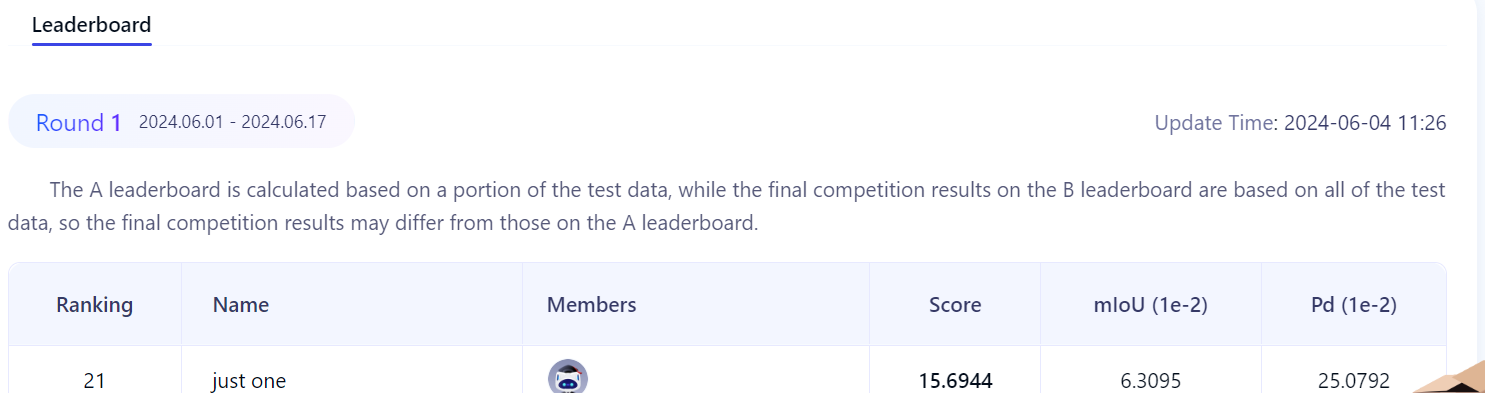
。

3.上午10:30所里组会内容记录

导师在外场实验，对我而言，没说啥，我交代了一下本周的计划。

4. 竞赛or平台搭建&跑代码【下午】 竞赛6.4今天要提交一次

竞赛，交了baseline。



遗留问题

1. 竞赛再改进看看。反正每天1次机会。别有压力。每天是死是活，瞎整1次。

0605

1.计算资源使用。平台搭建。

。

什么是xx

。

啥是xxx

。

遗留问题

？

0606-0607

1. xx算法

[https://www.](https://www.bilibili.com/video/BV19L4y1V7oG/?share_source=copy_web&vd_source=00215419ac39ea35bf3b245d1da97537)

0.来源论文：

1.应用：

2.背景：

3.相关论文：

4.解决的问题：

5.笔记截图：

6.背景知识：

①：

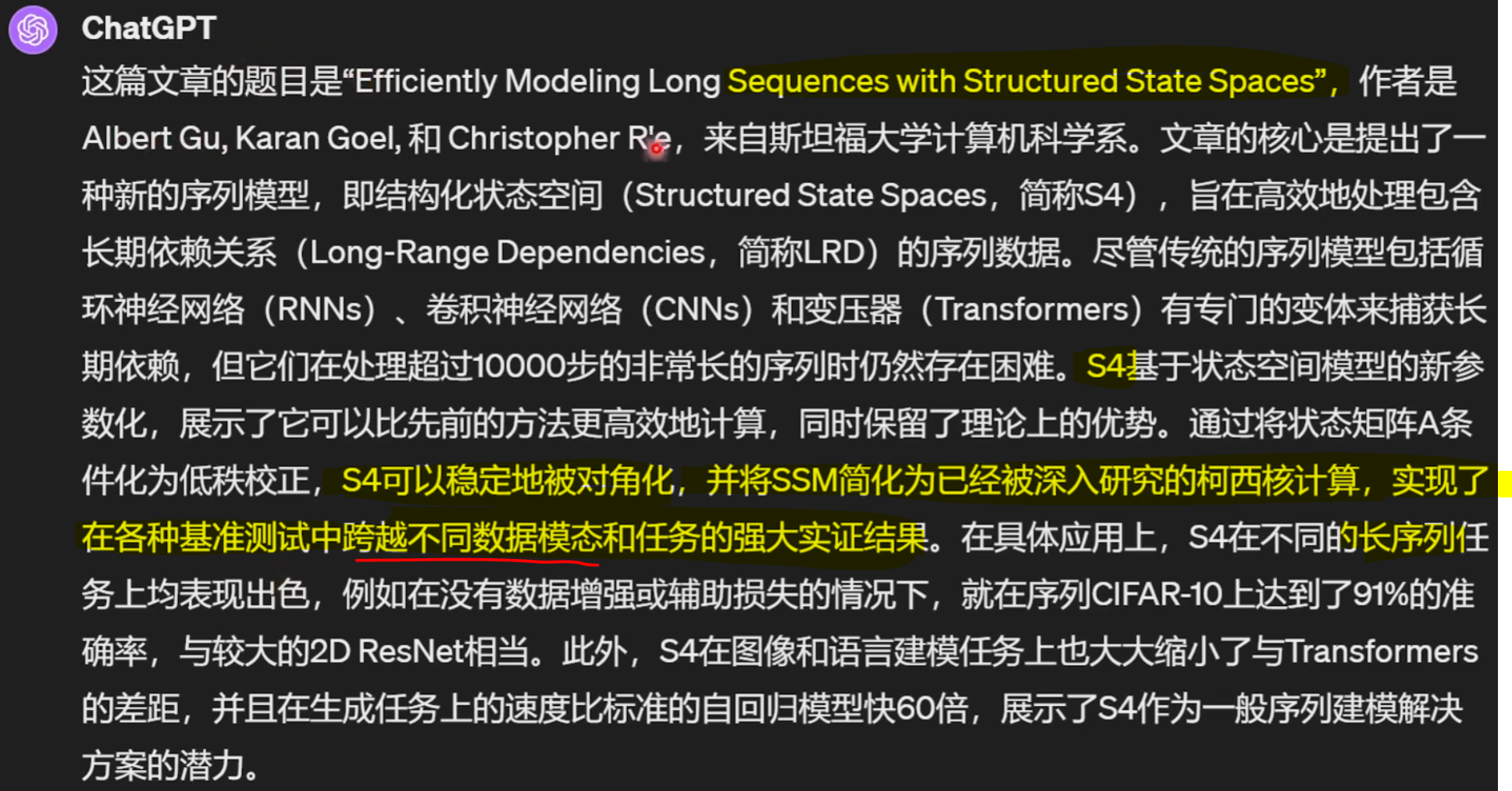
②：

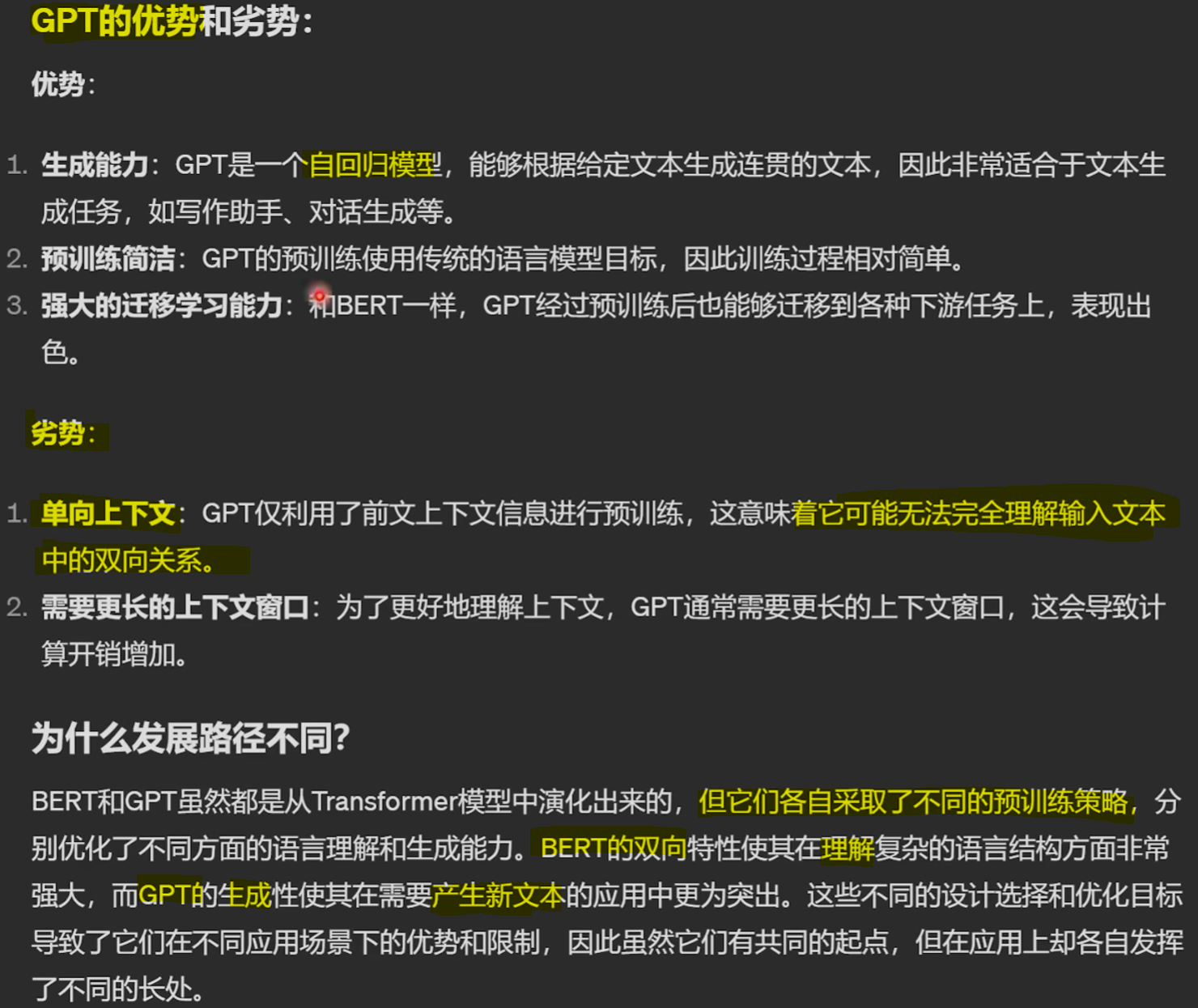
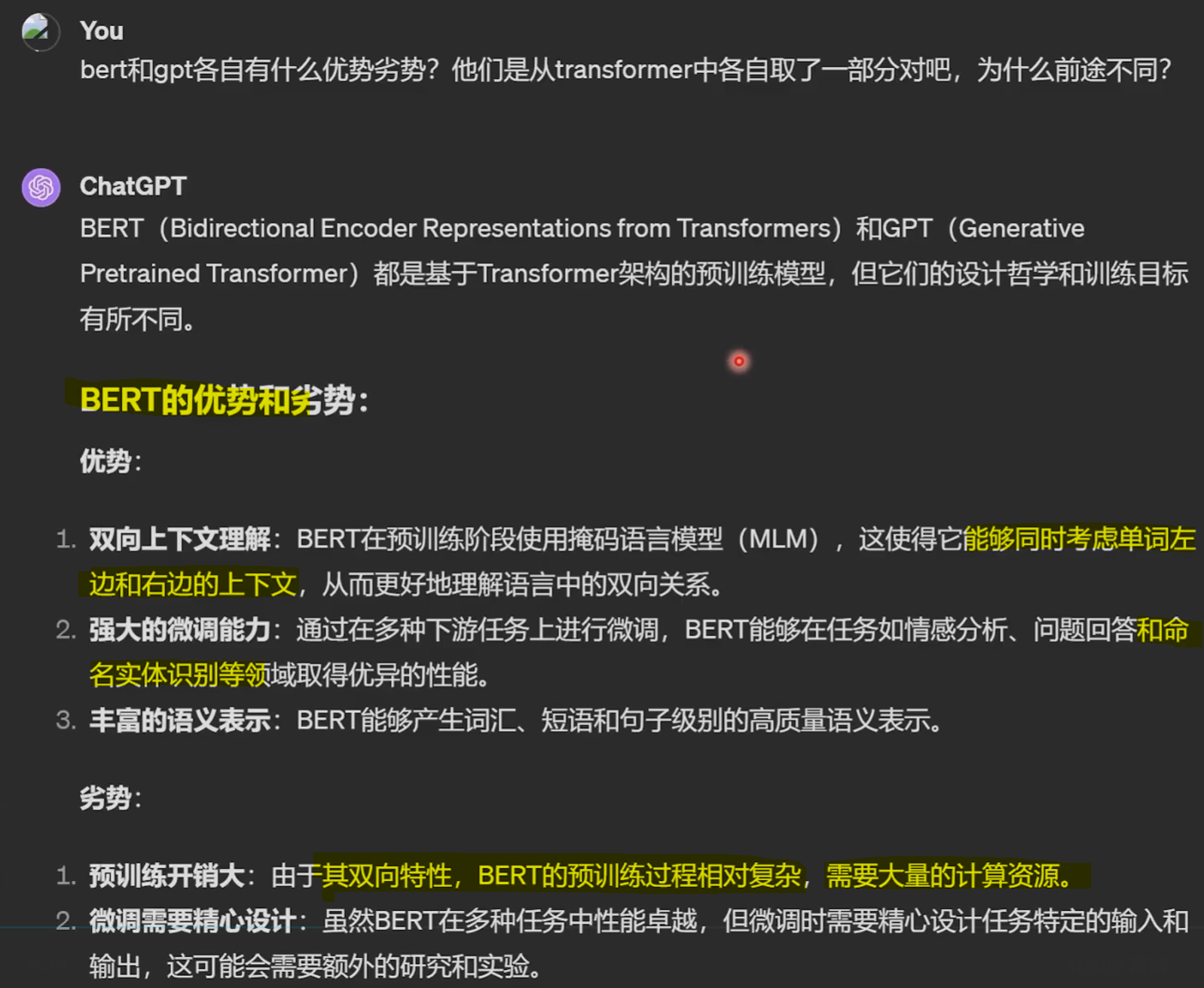
③xx不可忽略的原因：

④xx公式如图右侧

算法思想：

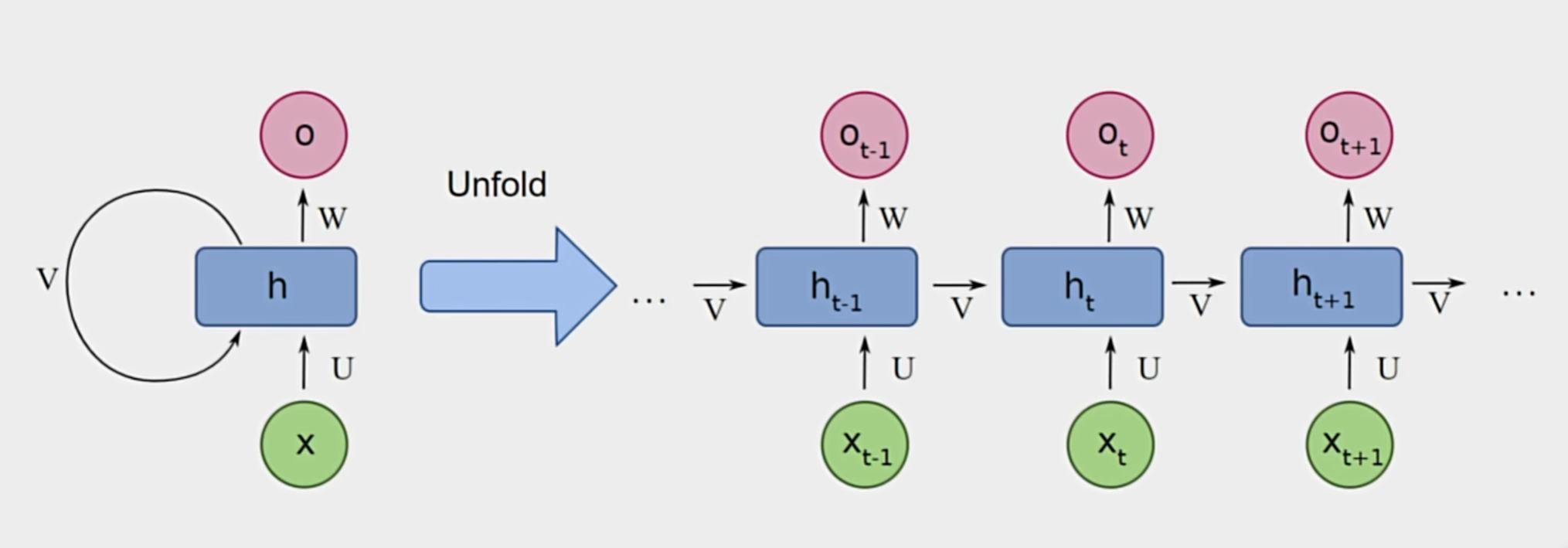
1. mamba之前的工作（同作者）





②Recurrent neural network（1986）

视频链接：<https://www.bilibili.com/video/BV1KH4y1W7cm/?spm_id_from=333.337.search-card.all.click&vd_source=1412c55689d7b78016c608ffdba64e83>



详细笔记截图

②

Xxxxx

详细笔记截图

③

Xxxxx

详细笔记截图

④

Xxxxx

详细笔记截图

遗留问题

Xxxxx