周报（2023. 5.9-2023.5.16）姓名：孙瑞阳

每日小结

	周一	周二	周三	周四	周五
早	修改网络代码，上课	大数据集去雾代码寻找	大数据集去雾，上课	多 loss 代码	阅读文献
中	多 loss 代码	论文阅读	多 loss 代码	遥感预训练 VIT 代码，上课	多 loss 代码
晚		上课	整理结果	遥感预训练 VIT 代码	遥感预训练 VIT 代码

注：简单表述当前时间段工作，如看文献 1，整理数据等

科研详情

文献阅读

文献1

题目： Align before Fuse: Vision and Language Representation Learning with Momentum Distillation

作者： Junnan Li, Ramprasaath R. Selvaraju, Akhilesh Deepak Gotmare, Shafiq Joty, Caiming Xiong, Steven Hoi

出处： arxiv 2021

方法：

三个模块：image encoder，text encoder，和 multimodal encoder，都用 transformer 建模，其中 multimodal encoder 每层多个 cross attention 来融合不同模态的信息。

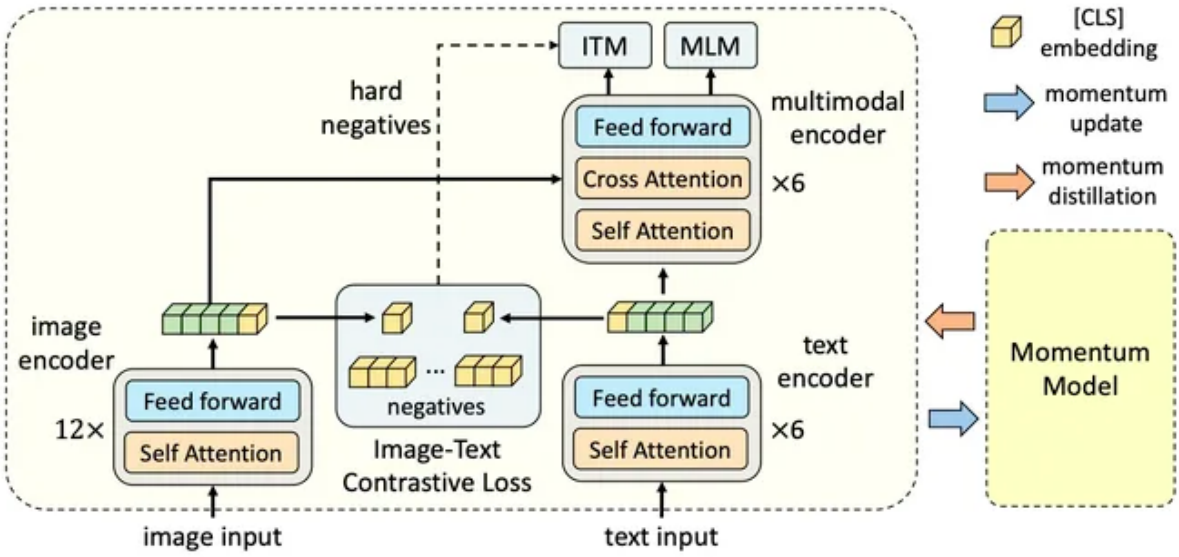


Figure 1: **Illustration of ALBEF.** It consists of an image encoder, a text encoder, and a multimodal encoder. We propose an image-text contrastive loss to align the unimodal representations of an image-text pair before fusion. An image-text matching loss (using in-batch hard negatives mined through contrastive similarity) and a masked-language-modeling loss are applied to learn multimodal interactions between image and text. In order to improve learning with noisy data, we generate pseudo-targets using the momentum model (a moving-average version of the base model) as additional supervision during training.

这里注意到 multimodal encoder 和 text encoder 其实参数量分别只有 image encoder 的一半，即它是一个 12 层的 transformer 劈开两半。

训练 loss：

ITC （Image-Text Contrastive Learning），image encoder 和 text encoder 分别对应的 cls token 的输出过个现行层，做对比学习。

该方法用 momentum networks 维护 memory bank 来用历史样本充当负样本：具体维护，image encoder 和 text encoder 各维护一个 momentum network，然后
online image encoder 和 momentum text encoder 做对比学习 online text encoder 和 momentum image encoder 做对比学习
MLM（Masked Language Modeling），随机 mask15%的文本 tokens，然后预测之。
ITM（Image-Text Matching），输入 image encoder 和 text encdoer 各个 token 的输出。其中，视觉的 tokens 输入到每一层的 cross attention，文本的 tokens 从底部输入 。最后文本的 cls token 的输出过个线性层预测文图是否匹配。
ITM 的正样本的 ITC 里的正样本，负样本则从 ITC 中选择最难的负样本。

MoD （Momentum Distillation）
动机：数据来源于网络噪声很大，类似 mean teacher 的方式，用 momentum network 来制作伪标签蒸馏。
作用模块：ITC 和 MLM
ITC：原来是 online image encoder 和 momentum text encoder，计算相似度，用 cross entorpy 训练；
这回，两边都用 momentum 的 encoder 计算相似度，然后用 KL 散度拉近两个的相似度分布（softmax 后的相似度向量），和原来的 itc loss 加权组合起来：

$$\mathcal{L}_{\text{itc}}^{\text{mod}} = (1-\alpha)\mathcal{L}_{\text{itc}} + \frac{\alpha}{2}\mathbb{E}_{(I,T)\sim D}\left[\text{KL}(\boldsymbol{q}^{\text{i2t}}(I) \parallel \boldsymbol{p}^{\text{i2t}}(I)) + \text{KL}(\boldsymbol{q}^{\text{t2i}}(T) \parallel \boldsymbol{p}^{\text{t2i}}(T))\right] \quad (6)$$

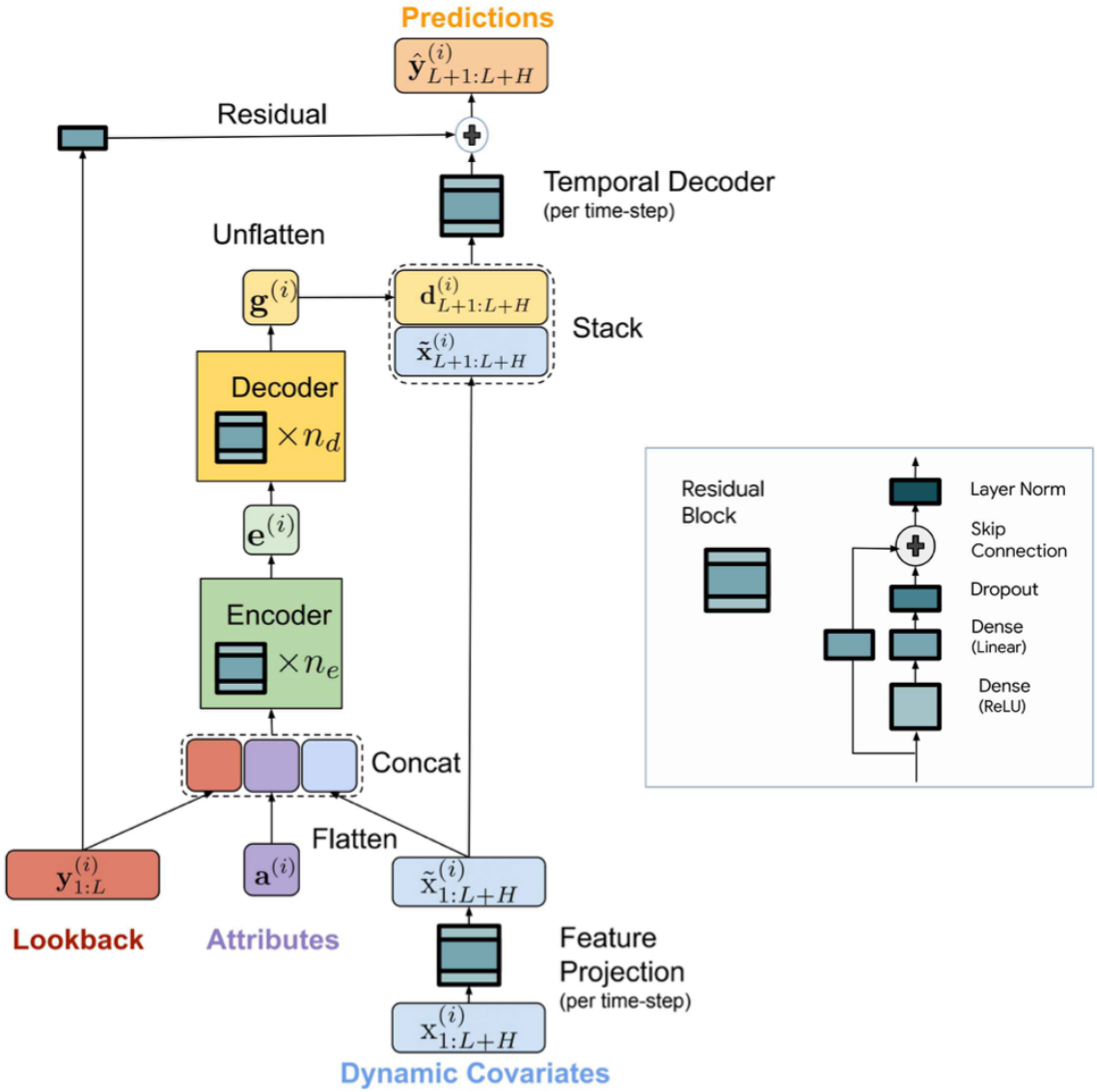
其中，红框部分为新增的蒸馏 loss。
MLM：用 momentum network，MLM 预测的结果作为 soft-label，用 KL 散度逼近之：

$$\mathcal{L}_{\text{mlm}}^{\text{mod}} = (1-\alpha)\mathcal{L}_{\text{mlm}} + \alpha\mathbb{E}_{(I,\hat{T})\sim D}\text{KL}(\boldsymbol{q}^{\text{msk}}(I,\hat{T}) \parallel \boldsymbol{p}^{\text{msk}}(I,\hat{T})) \quad (7)$$

启发：

1. Align Before fuse：用对比学习 loss 把图像、文本数据的 embedidng 对齐，然后把图像、文本 embedding 融合起来做其他任务（ITM 和 MLM），凑齐 VLP 同时训练。 **在看代码，想增加这个 loss 在自己的工作中**
2. 用 Momentum distillation 来克服 noisy data，即用 momentum Network 来生成伪标签，作用在 ITC 和 MLM 上，甚至在下游任务上。
3. 缺点很明显，做 N 多个任务，多个网络，一次迭代要前传很多次。

文献2
题目：Long-term Forecasting with TiDE: Time-series Dense Encoder
作者：Abhimanyu Das, Weihao Kong, Andrew Leach, Rajat Sen, and Rose Yu
出处：arxiv 2023
方法：
本文提出了 TiDE 模型，整个模型没有任何注意力机制、RNN 或 CNN，完全由全连接组成。实验中 TiDE 效果超越了各个 Transformer 时间序列预测模型（PatchTST、FEDformer、Auotoformer、Informer 等）
模型的核心基础组件是 Residual Block，由一个 Dense+RLU 层、一个 Dense 线性层、一个 Add&Layernorm 组成。TiDE 其他组件都基于这个基础 block 搭建。
模型整体可以分为 **Feature Projection、Dense Encoder、Dense Decoder、Temporal Decoder** 四个部分。



Feature Projection 将外部变量映射到一个低维向量，使用 Residual Block 实现，主要目的是对外部变量进行降维。

Dense Encoder 部分将历史序列、属性信息、外部变量映射的低维向量拼接到一起，使用多层 Residual Block 对其进行映射，最终得到一个编码结果 e。

Dense Decoder 部分将 e 使用同样的多层 Residual Block 映射成 g，并将 g 进行 reshape 成一个[p, H]的矩阵。其中 H 对应的是预测窗口的长度，p 是 Decoder 输出维度，相当于预测窗口每个时刻都得到一个向量。

Temporal Decoder 将上一步的 g 和外部变量 x 按照时间维度拼接到一起，使用一个 Residual Block 进行每个时刻的输出结果映射，后续会加入历史序列的直接映射结果做一个残差连接，得到最终的预测结果。

启发：

1. 整个模型没有任何注意力机制、RNN 或 CNN，完全由全连接组成。**在时序数据上，可能不是越复杂的模型效果就越好，**时序的本质是多个不同周期的傅里叶级数，和未能预测的要素体现为噪音，二者的叠加。 算法好与否，就看能不能从观察数据中，分离周期和噪音，进一步如何能把多周期提取出来并对应到其实际物理意义。

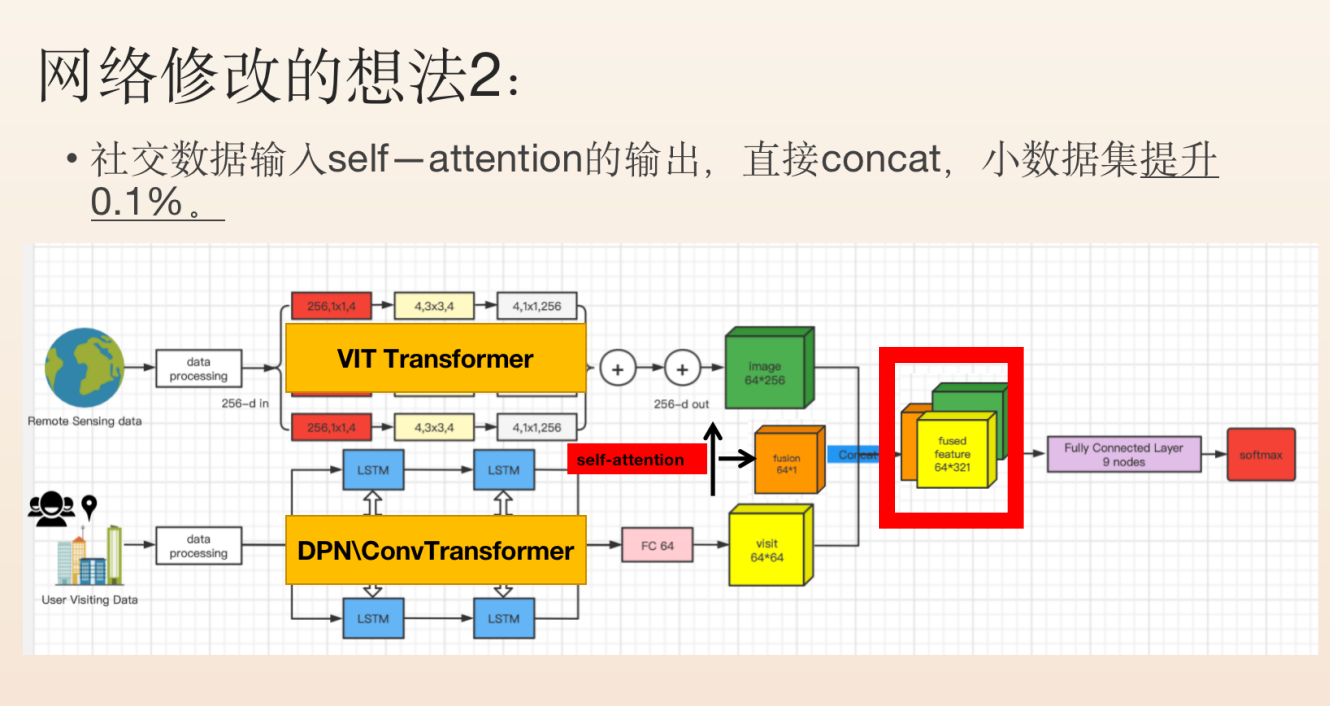
工作进展

1：阅读文献；

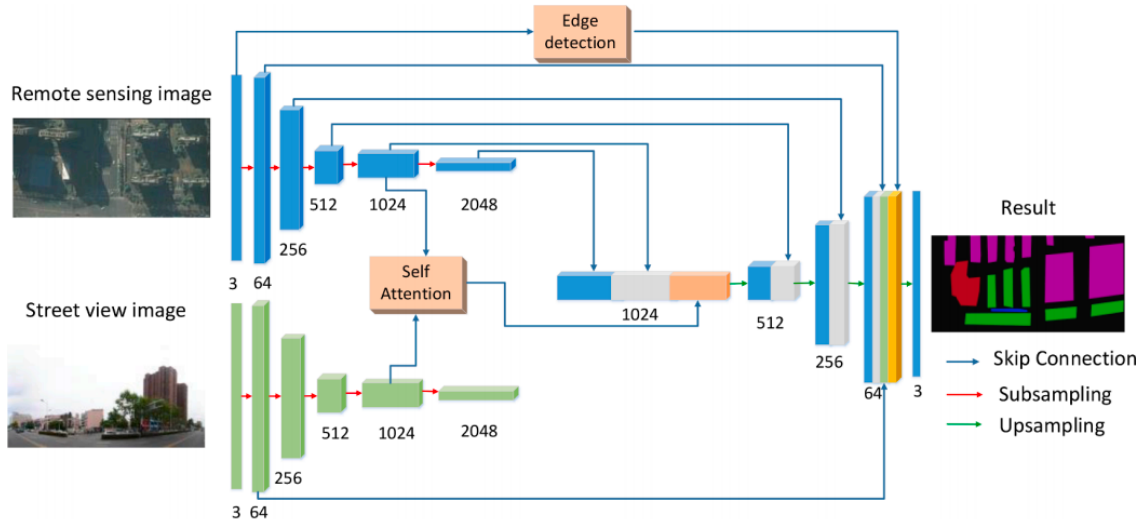
2：补充了大小数据集上的实验

3. 在原有交叉熵loss上增加loss：**Balanced Cross Entropy代码已跑通（跑的很慢，loss太大了有点训练），Focal loss已跑通，多loss平衡正在学习**

4. 增加了修改网络的想法2，小数据集提升0.1%：



该想法来自于论文：Flood vulnerability assessment of urban buildings based on integrating high-resolution remote sensing and street view images



5. 大数据集增加去雾预处理，在改 bug。

6. 遥感预训练大数据集VIT，在套代码，还在改bug（有点难改）。

下周计划

1. 大数据集增加去雾预处理
2. 网络修改继续看论文想 idea
3. 遥感预训练大数据集VIT尽快改完bug
4. **在原来的交叉熵loss上增加了Balanced Cross Entropy、 Focal loss，多loss平衡正在学习**