周报（2023. 3.21-2023.3.28）姓名：孙瑞阳

每日小结

	周一	周二	周三	周四	周五
早	Autoformer 代码，上课	Autoformer 代码，上课	大数据集处理，上课	上课，大数据集处理	Timenet 代码
中	Autoformer 代码，上课	论文阅读	整理代码	大数据集处理，上课	Timenet 代码
晚	上课，Autoformer	组会	大数据集处理	大数据集处理	跑 Timenet

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

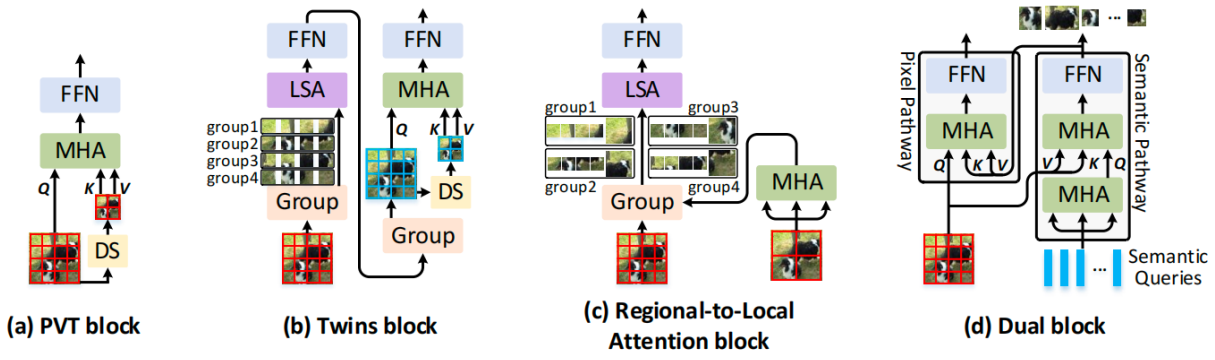
题目： Dual Vision Transformer

作者： Ting Yao, Yehao Li, Yingwei Pan , Yu Wang, Xiao-Ping Zhang , and Tao Mei,

出处： TPAMI 2023

方法：

作者提出了一种新的 Transformer 结构，即双视觉 Transformer（双 ViT）。本文的出发点是使用特定的双通道设计升级典型的 Transformer 结构，并触发全局语义和局部特征之间的依赖关系，以增强自注意力学习。

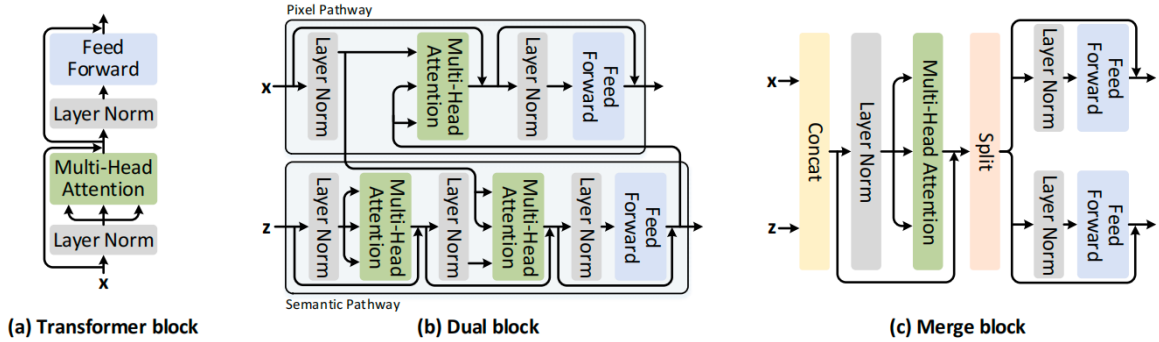


具体而言，双 ViT 由四个阶段组成，其中每个阶段的特征图分辨率逐渐缩小。在具有高分辨率输入的前两个阶段中，双 ViT 采用了新的双块，由两个路径组成：

（i） 像素路径，通过在像素级重新定义输入特征来捕获细粒度信息

（ii） 语义路径，在全局级抽象高级语义 token。语义路径稍深（操作较多）

但从像素中提取的语义 token 较少，像素路径将这些全局语义视为在学习较低像素级细节之前的语义。这种设计方便地编码了内部信息对整体语义的依赖性，同时降低了高分辨率输入下多头自注意力的计算成本。在最后两个阶段，这两条路径的输出被合并在一起，并进一步反馈到多头自注意力中。



Dual Block

作者设计了一个针对高分辨率输入（即前两个阶段）的原则性自注意力块，即双块。新的设计很好地引入了一个额外的途径来缓解自注意力学习。上图（b）描述了双块的详细架构。具体来说，双块包含两条路径：像素路径和语义路径。语义路径将输入特征映射总结为语义 token。之后，像素路径以键/值的形式优先考虑这些语义 token，并通过交叉注意力对定义的输入特征图进行多头注意力。

Merge Block：

前两个阶段中的双块利用了两条路径之间的相互作用，同时由于高分辨率输入的巨大复杂性，像素路径中的局部 token 之间的内部相互作用未被利用。为了缓解这个问题，作者提出了一种简单而有效的自注意力块（即合并块）设计，以在最后两个阶段（使用低分辨率输入）对 concat 的语义和局部 token 执行自注意力，从而实现局部 token 之间的内部交互。

启发：

1. 考虑到梯度通过语义和像素路径反向传播，DUAL 块能够同时通过像素到语义的交互来补偿全局特征压缩中的信息损失，并通过语义到像素的交互来减小局部特征提取与全局先验的差异。

文献2

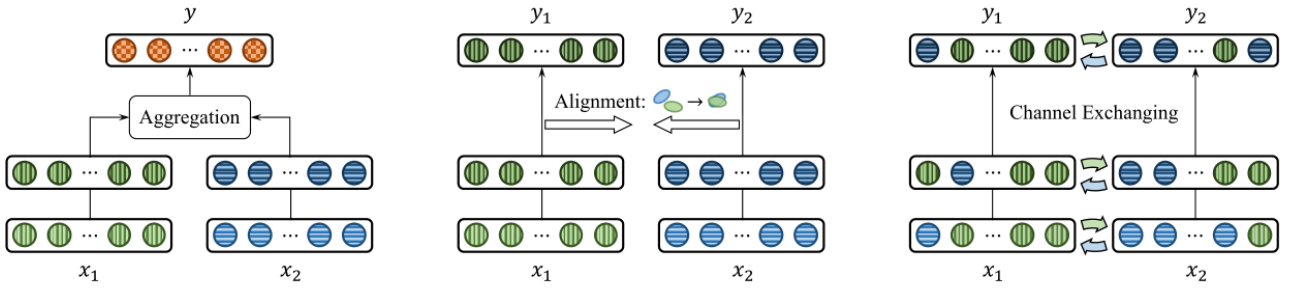
题目：Deep Multimodal Fusion by Channel Exchanging

作者：Yikai Wang, Fuchun Sun, Wenbing Huang, Fengxiang He, Dacheng Tao

出处：TPAMI-2022

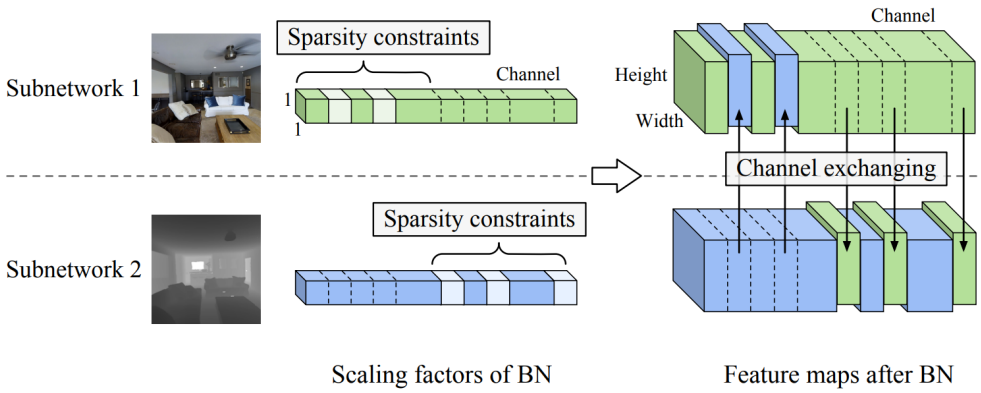
方法：

本文提出了通道交换网络（CEN），一个无参数的多模态融合框架，在不同模态的子网络之间动态地交换通道。具体来说，信道交换过程是由单个信道的重要性自我引导的，这个重要性是由训练期间的批量标准化（BN）缩放因子的大小来衡量的。这种交换过程的有效性也是通过共享卷积滤波器，但在不同的模式下保持独立的 BN 层来保证的



(a) Aggregation-based fusion **(b)** Alignment-based fusion **(c)** Ours

CEN 自适应地在子网络之间交换信道。**CEN 的核心在于其受网络修剪启发的较小范数信息量较少的假设。 具体而言，利用批量归一化（BN）或实例归一化（IN）的缩放因子（即 γ）作为每个相应信道的重要性度量，并用其他子网络的平均值替换与每个子网络的接近零因子相关的信道。CEN 的另一个特点是，除了所有子网络的 BN 层之外，参数是彼此共享的。**



Scaling factors of BN Feature maps after BN

启发：

1. 文章公式多，代码较为简单，还在一边看代码一边看公式，看能否套用在特征融合部分

工作进展

1：阅读文献；

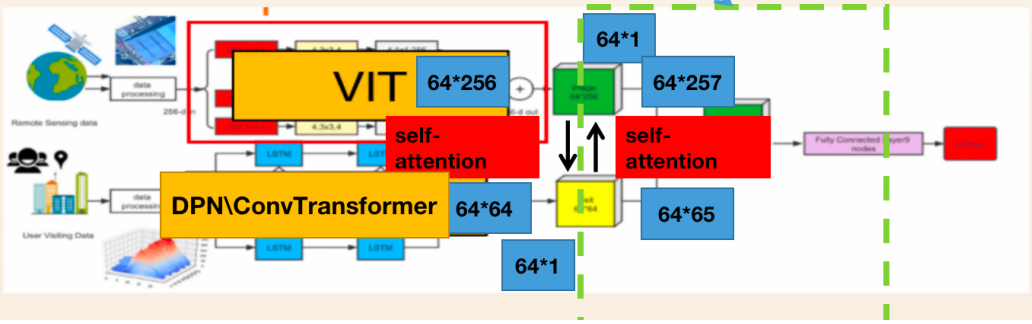
2：补充了大小数据集上的实验：

网络修改的想法：self—attention

- 遥感特征输入self—attention输出并concat到社交数据特征。
- 社交数据输入self—attention输出并concat到遥感数据特征。（已经实现，像素级64*1）

达到两者特征在**通道或者像素级上**的选择融合

此处是否可以换成多头注意力，或者其他结构？



网络修改self—attention效果：基本都提升了0.2-0.5%

- 社交数据输入self—attention输出并concat到遥感数据特征。（已经实现，像素级64*1）
- 小数据集：

	小数据集					
	网络修改，增加双分支间的self-attention					
	VIT+DPN	VIT+ConvTransformer(1*5卷积核),head=8	VIT+ConvTransformer(1*24卷积核),head=8	VIT+ConvTransformer(1*24卷积核,head=1)	VIT+ConvTransformer(都是1*24卷积核),head=8	Muti-scale VIT transformer+Conv（都24）
无self-attention	0.694	0.675	0.676		0.67	
有从社交输入到遥感的self-attention	0.687	0.677	0.678	0.677	0.675	0.666

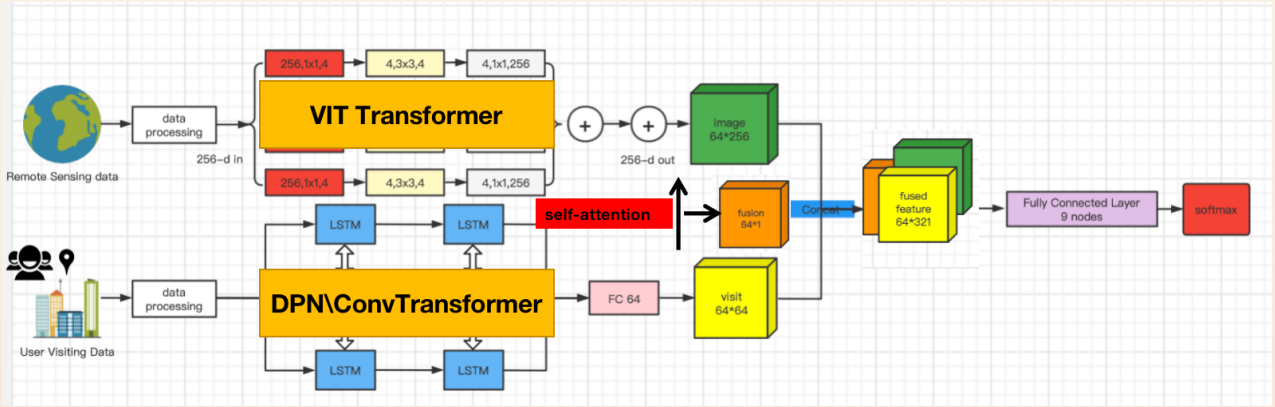
- 大数据集

	大数据集			
	网络修改，增加双分支间的self-attention			
大数据集	VIT+DPN	VIT+ConvTransformer(1*5卷积核),head=8	VIT+ConvTransformer(1*24卷积核),head=8	VIT+ConvTransformer(1*24卷积核,head=1)
无self-attention	0.667	0.671	0.6705	
有从社交输入到遥感的self-attention	0.666\0.549	0.675	0.676	

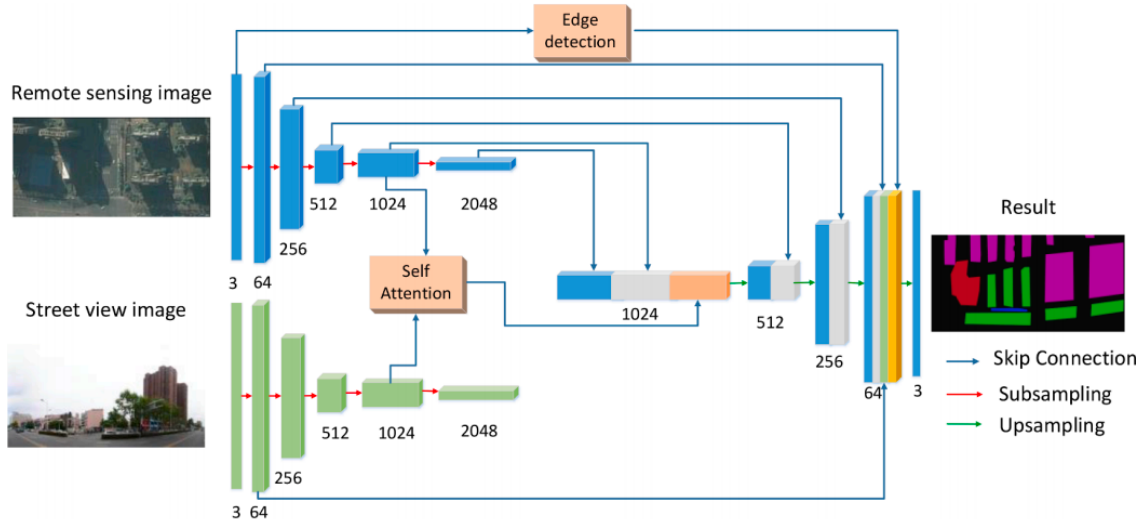
3. 增加了修改网络的想法，在跑，有提升，还没跑完：

网络修改的想法：

- 社交数据输入self—attention的输出，直接concat，还没跑完，有提升。



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



4. 遥感预训练大数据集VIT，在套代码，改bug。

下周计划

1. 阅读文献
2. 大数据集增加去雾预处理
3. 增加 loss
4. 网络修改继续看论文想 idea
5. 遥感预训练大数据集VIT尽快改完bug