# 周报（2023. 6.6-2023.6.13）姓名：孙瑞阳

## 每日小结

| | 周一 | 周二 | 周三 | 周四 | 周五 |
|---|---|---|---|---|---|
| 早 | VALSE，跑代码 | PPT，跑代码，整理结果 | 分支间多头注意力代码 | 遥感楼检修电路，服务器关了，<br>读文献想 loss | VALSE，跑代码 |
| 中 | Loss 想法代码 3，<br>VALSE，跑代码 | 论文阅读 | 复现 isprs 代码 | 遥感楼检修电路，服务器关了，<br>读文献想 loss | VALSE，跑代码 |
| 晚 | 实验补充，ppt | | 整理结果 | | 双重 Fusion 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**

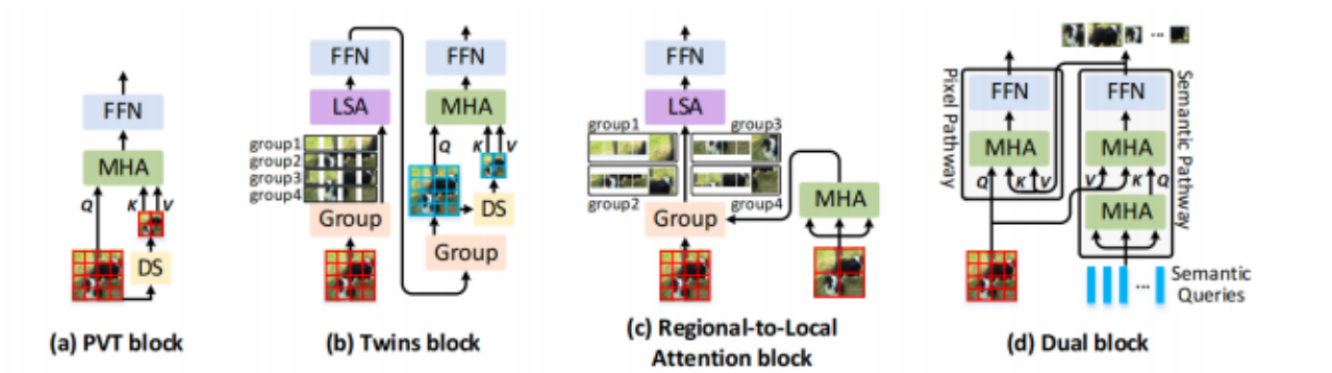
**题目：** Dual Vision Transformer

**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。

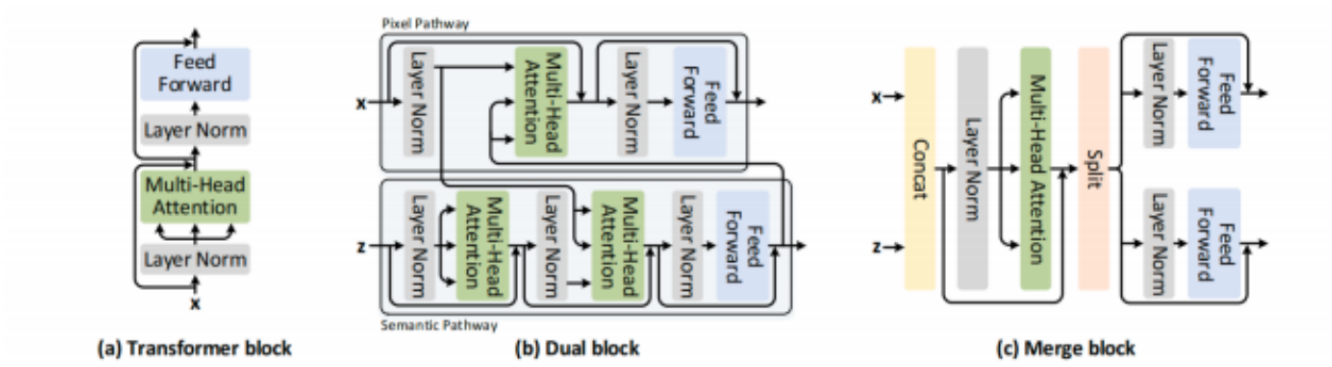



具体而言，双 ViT 由四个阶段组成，其中每个阶段的特征图分辨率逐渐缩小。在具有高分辨率输入的前两个阶段中，双 ViT 采用了新的双块，由两个路径组成：

(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。

2.本文的 Dual Block 中用的是多头注意力，可以尝试更换我目前的 self attention，在看代码。

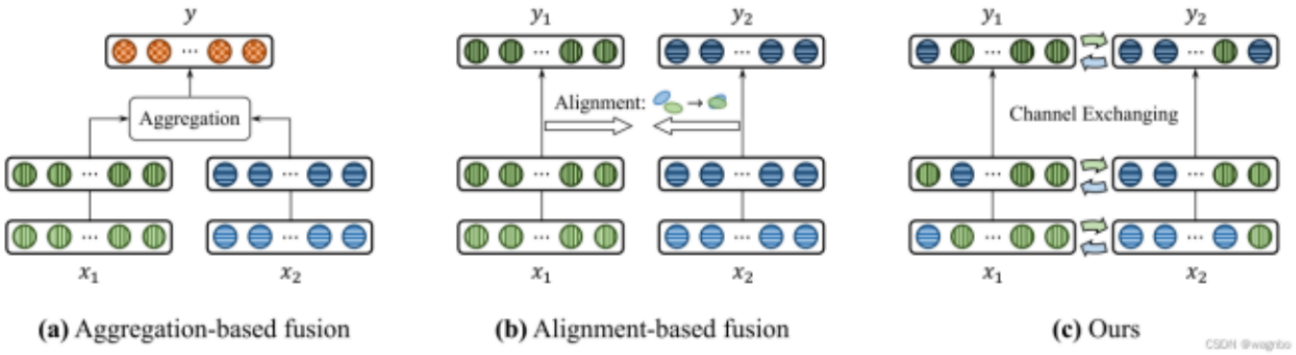
**文献2**

**题目：Deep Multimodal Fusion by Channel Exchanging**

**作者： Yikai Wang, Fuchun Sun, Wenbing Huang, Fengxiang He, Dacheng Tao 出处：TPAMI-2022**

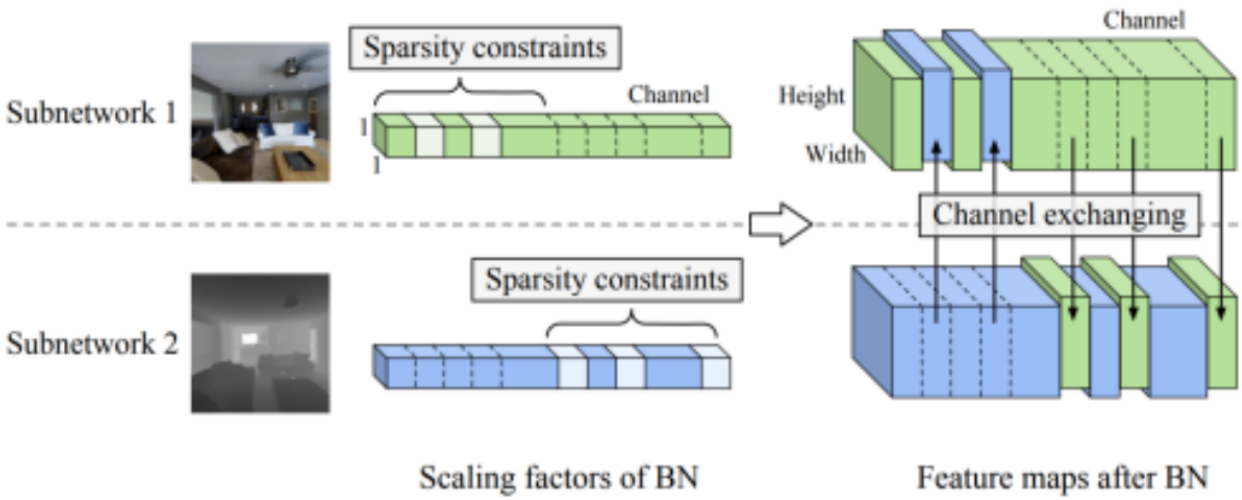
**方法：**

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




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 $\gamma$)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。




**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

**VALSE：**

**重庆大学 张磊**

介绍迁移学习/领域自适应的概念基础和过去 10 年来最新研究进展。此外，介绍他们的研究团队近年来在 TL / DA 方面的研究工作，包括分类器自适应模型（EDA，TIP16），子空间重建迁移学习（LSDT，TIP16; CRTL，ICCV17），公共子空间迁移学习（CDSL，TIM17），流形准则迁移模型（MCTL，TNNLS18）和自我对抗迁移网络（AdvNet，ACM MM17）。

启发：

进行多模态研究时，一般从两个角度：

学习度量，使得两个模态的数据信息更加相似，学习出两个模态共同的特征。比如可见光与近红外人脸识别。

2.就是结合多模态的特征信息，将其 modality-specific 的特征进行级联，提升识别效果。

**武汉大学 武宇**

复杂场景下的多模态信息感知与生成：本报告从理解和生成两个角度对现有的多模态模型进行分析。首先介绍了视觉-语言指代检测和分割任务的前沿现状，并指出绝大部分提升来自于额外大数据上的预训练过程。通过对模型内部热力图的分析，报告指出跨模态编码器是已有模型的瓶颈，并基于此提出了阶段性权重重置策略来不断激活跨模态编码器的训练，从而实现了小模型能达到和接近预训练后的大模型的效果。在此基础上，武宇教授介绍了其课题组在扩散模型 AIGC 方向的最新工作，包括通过对比学习增强扩散模型输入和输出的互信息，以及使用通过对扩散模型进行反演实现任意属性任意强度的图像编辑。

**中科院自动化所 刘静**

多模态预训练模型的研究与应用：回答了多模态预训练（1）为什么关注？（2）当前怎么做？（3）以后怎么做？等三个问题。指出多模态大模型如 ChatGPT 之所以被广泛关注是由于其通用的意图理解能力、强大的连续对话能力、智能的交互修正能力和较强的逻辑推理能力。在对预训练模型的核心思想进行介绍的基础上，大模型从单模态迈向多模态成为必然，并从多模态预训练数据集、基础模型、面向多模态下游任务的模型微调等多个方面对模型性能国际领先的大模型进行了介绍。

**南京理工大学 李泽超**
汇报了开放环境下细粒度多媒体内容分析与检索问题，主要是弱监督视觉细粒度分析推理、基于语义遮挡的跨模态检索、基于深度协同因子分解的多模态内容检索、基于区域定位哈希的细粒度图像检索

**北京大学 张健**
提出了去噪扩散零空间模型(DDNM)，DDNM 只需要一个预先训练好的现成扩散模型作为生成先验，不需要任何额外的训练或网络修改。通过对反向扩散过程中的零空间内容进行细化，可以得到既满足数据一致性又满足数据真实性的多种结果。进一步提出了一种增强的鲁棒版本，称为 DDNM+，以支持噪声恢复并提高任务的恢复质量。

**中国科学院自动化研究所 张兆翔**
从 2D 视觉场景检测技术出发，探讨了基于 3D 空间信息以及 4D 时空信息的视觉场景感知方案，构建了通用的 2D 目标检测框架、高效 3D 目标检测方案以及无监督的 4D 目标检测框架等。特别的是融入了时序网络 FEDformer 框架进行时序信息下的目标检测。

**浙江大学 朱霖潮**
详细讨论了基于基于纯文本的零样本描述生成方法。零样本描述生成技术通过训练文本解码器与无参数的映射单元，使得测试时通过图片内容即可自动生成文本描述。
介绍了基于提示学习的多模态融合方法，指出面向计算受限场景中，需要优化提示学习方法，迭代式融合多模态信息。模型中的多种知识存在鸿沟，从整体上对齐模态对跨媒体理解有重要作用，以及多重信息融合应该从内存、参数量、精度等多方面衡量融合效果。未来可重点研究提示词的可解释性、异构知识融合、新型的结构化多模态表达和数据与知识联合建模框架等。

工作进展

1：阅读文献；

2： 整理近五年相关工作

3: 补充了单模态的小数据集实验：多模态分别提高了 29.3%和 9.5%，在绘制混淆矩阵：

| Data Mode | Accuracy |
|---|---|
| Remote Sensing data | 39.3% |
| Social Perception data | 59.1% |
| Two modes data | 68.6% |

**原论文里，小数据集最好的是 70.02%，我现在是 68.6%；大数据集最好他是 70.23%，我是 67.8%。**

4:增加深度监督的 loss 以后，小数据集又提了 0.5 个点，增加后有 0.686，大数据集还在跑

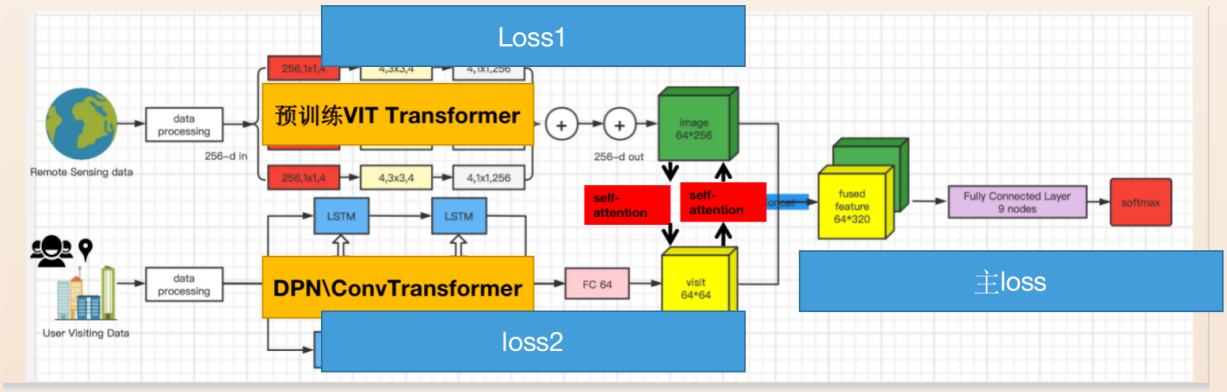
自己的想法：

分支loss和主loss的引导：

小数据集：0.686，涨0.5%，目前best

大数据集：没跑完

- 三个loss都使用交叉熵：Loss=主Loss+a(Loss1+Loss2)

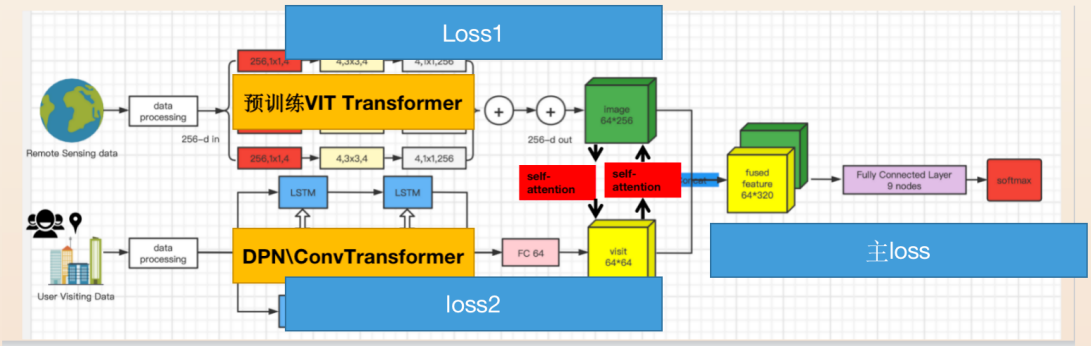



5.增加深度蒸馏的 loss 后，小数据集有提升，还在调参，感觉能更好

自己的想法：

拿预测真值去反向传播两个小loss，分支loss和主loss的引导（小数据集有提升，还在调，感觉能更好）

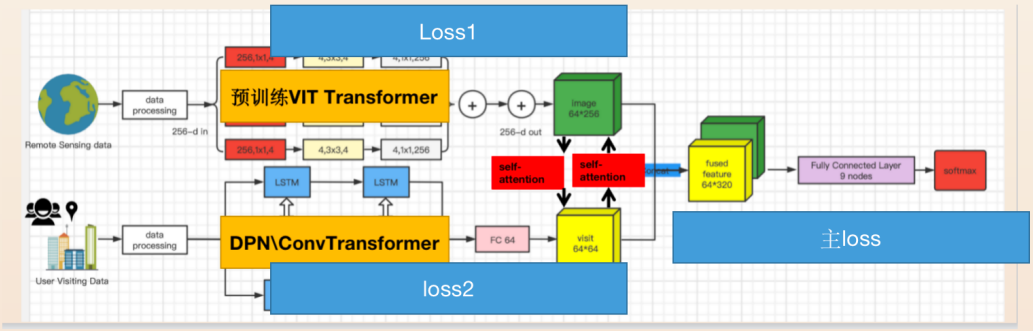
- 三个loss都使用交叉熵：Loss=主Loss+0.2(Loss1+Loss2)




6.同时在交叉熵的基础上增加深度监督的 loss 和深度蒸馏的 loss，小数据集有提升，还在调参，感觉能更好

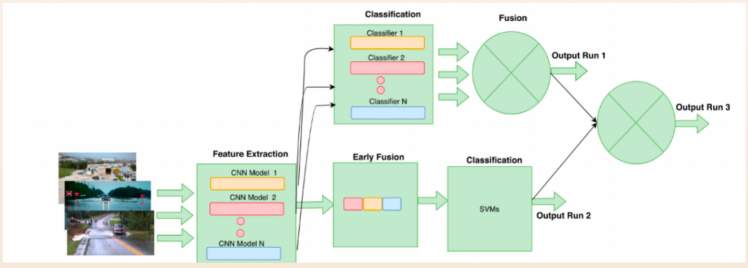
自己的想法：监督loss+蒸馏loss+交叉熵
（小数据集有提升，还在调参，感觉能更好）

- 三个loss都使用交叉熵：Loss=主Loss+a*监督loss+b*蒸馏loss




7.双重融合，效果很差：

双重融合，效果不好，掉了十个点




8.多头注意力代码，在改 bug
9.遥感何恺明多尺度代码，有点麻烦，还没时间改 bug

下周计划

1. 网络修改：继续看论文想 idea：多头注意力代码，遥感何恺明多尺度代码
2. 撰写摘要、introduction
3. **准备20号组会ppt，组会还需要补充很多实验，这周需要跑很多代码，已经找苏老师开了超算，服务器也不够用。。。**
4. 类似双重融合，用花里胡哨的简单fusion方法修改一下fusion的concat，**还想看看变化检测里做的一些简单花里胡哨的fusion，增加创新点**