

### 每日小结

	周一	周二	周三	周四	周五
早	论文阅读	上课	上课	阅读文献	上课
中	上课	论文阅读，上课	论文阅读	上课	上课，模型学习
晚	新生讲课准备	组会	新生讲课	上课	上课

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

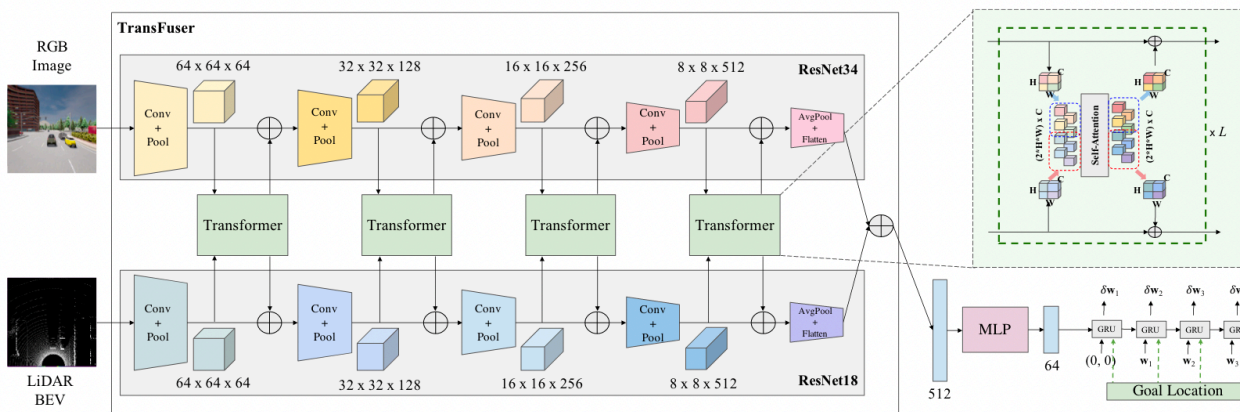
题目：Memory Fusion Network for Multi-View Sequential Learning

作者：Amir Zadeh, Paul Pu Liang, Navonil Mazumder, Soujanya Poria, Erik Cambria, Louis-Philippe Morency

出处：CVPR 2021

方法：

在该 TransFuser 模型中，将单视图 RGB 图像和 LiDAR 的 BEV 作为多模态融合转换器 (TransFuser) 的输入，该转换器使用多个转换器模块来融合两种模态之间的中间特征图。这种融合在整个特征提取器中以多种分辨率 ( $64 \times 64$ 、 $32 \times 32$ 、 $16 \times 16$  和  $8 \times 8$ ) 应用，最终图像和 LiDAR BEV 流中输出 512 维特征向量，通过组合逐元素求和。输出的 512 维的特征向量构成了对 3D 场景的全局上下文编码的环境表达。接下来使用 MLP 对其进行处理。在 MLP 模块中，采用一个单层 GRU 和线性层的结构，该层采用隐藏状态预测车辆航路点。模型输出以自身车为中心，一系列时间内的路径点。

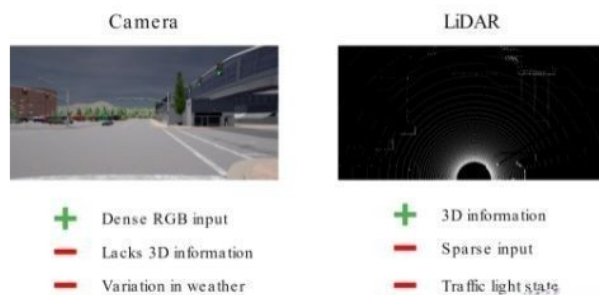


解决问题：

在这里，相机可能看得到交通灯，激光雷达可以看到红色区域的车，所以在将两者结合后 自身就知道不应该在此时立刻右转。也就是对自身的整个 3D 场景进行建模。



典型场景



camera和LiDAR各自优缺点

启发:

1. Transfuser 在特征提取器中以多种尺度 ( $64 \times 64$ 、 $32 \times 32$ 、 $16 \times 16$  和  $8 \times 8$ ) 应用, 并通过消融实验证明多尺度融合可以有效地整合来自不同模态的特征。考虑自己的工作, 如何在遥感图像多尺度处理, 目前还在看这篇论文的代码。
2. 模型输入, 考虑自己的工作, 如何在社交数据处理维度, 能够较好的对齐:
  - (1) LiDAR 的 BEV ( $256 \times 256 \times 2$ ): 车辆前方 32m, 左右 16m 范围点云加入考虑, 得到  $32 \times 32 \times 2$ , 以 0.125 作为分辨率, 整个像素值为  $256 \times 256$
  - (2) 单视图 RGB 图像 ( $256 \times 256 \times 3$ ): 相机 FOV100 度, 收到 RGB 是  $400 \times 300$ , 转成  $256 \times 256$
3. 代码正在学习, 并复现原有数据集。

## 文献2

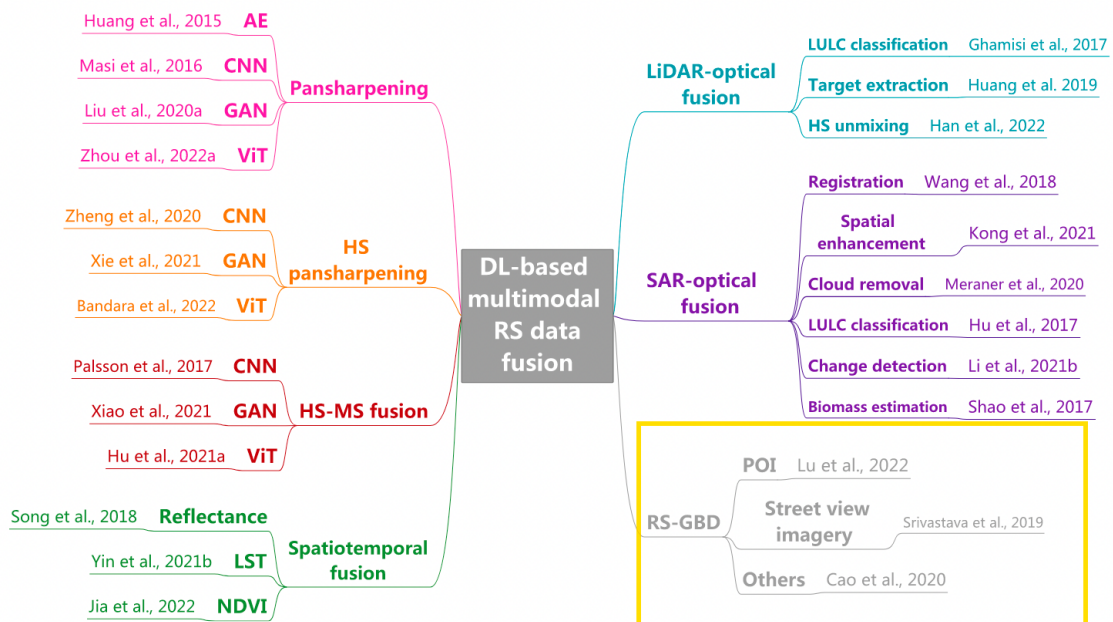
题目: Deep learning in multimodal remote sensing data fusion: A comprehensive review

作者: Jiaxin Li, Danfeng Hong, Lianru Gao, Jing Yao, Ke Zheng, Bing Zhang, Jocelyn Chanussot

出处: International Journal of Applied Earth Observation and Geoinformation

方法:

该综述根据然后根据待融合的数据模态将相关研究进行分类: 空间光谱、时空、光检测和测距-光学、合成孔径雷达-光学和 RS-地理空间大数据融合。



启发:

1. 读了其中一篇《A unified deep learning framework for urban functional zone extraction based on multi-source heterogeneous data》, 做 POI 转层次矩阵热图和遥感数据使用 CNN 和 attention 融合的论文, 找到了代码和数据, 还在读。
2. 找到了一个公开的腾讯用户密度数据和代码, 还在研究数据。

## 工作进展

- 1: 学习《Multi-Modal Fusion Transformer for End-to-End Autonomous Driving》代码, 有难度, 还在学习, 尝试复现;
- 2: 复现了两个集成学习lightGBM和XGBoost融合社交和遥感数据的代码;
- 3: 新生讲课: 准备新生讲课讲稿, 修改ppt。