**AGT International**

AGT international 是2017年IEEE数据融合大赛 (IEEE Data Fusion Contest, DFC) 第二名组。

**内容：**AGT通过多种方法集成的策略对目标地区多种数据进行融合，并实现LCZ分类，集成方法包括随机森林 (Random Forest, RF)、CNN和梯度提升机（Gradient Boosting Machine, GBM），选择三种方法的原因：RF具有较好的鲁棒性和稳定性，GBM不容易过拟合，CNN能够有效获取空间关系。

**问题与挑战：**

1. 对深度学习来说，样本量太少（需要进行数据扩张）
2. 训练样本类别数量不均衡（部分样本数量只有300）
3. 数据质量问题（噪声，数据获取时间，云量等）
4. 不同城市影像数量不同
5. 必须进行数据融合才能完成识别任务，而不能只利用光谱数据
6. 光谱数据与OSM数据的分辨率不同

**流程：**

**1** 数据丰富：OpenStreetMap(OSM)数据；从EarthExplorer获取的额外的影像数据。（这里不太理解，附1）

**2** 特征提取：

OSM：土地利用信息；及从OSM中统计的建筑物特征（研究区的建筑物数量、建筑物的平均及最大层数、建筑物的平均及最大高度）。

光谱特征：多种光谱特征（NDVI, NDWI, BSI, Normalized difference moisture index, SAM, MNF），由于每个地点有一幅Sentinel-2影像和两幅Landsat 8影像可用，因此本文将Sential-2数据叠加至每幅Landsat 8数据上。

**3** 特征筛选：使用随机森林法对以上特征进行筛选，随机生成一个符合[0, 1]上均匀分布的特征，如果目标特征的重要性大于随机特征的重要性，则保留此特征作为分类器训练用的特征。

**4** 分类器：

CNN：一个只包含三层的浅层网络架构。第一层是一个批标准化 (Batch normalization)的过程，第二层是卷积层，第三层是Softmax层，计算每个LCZ分类的后验概率。（输入数据为原始多光谱影像）

RF：本文使用了一种加权RF方法，即根据训练样本中类别频率进行反向加权，以解决类别数量不均衡的问题。（输入数据为所有人工挑选的特征）

GBM：进行了超参数调优。（输入数据为所有人工挑选的特征）

**5**方法集成：基于最小熵的集成方法。三层：第一层将同一地点的分类结果中质量较高的结果，第二层在第一层的基础上将五个模型结果进行聚合，第三层在第二层基础上将三个时间的结果进行聚合，最后将三种分类方法结果进行结合。

**6** 结果平滑：采用马尔可夫随机场模型。

**结果：**总体精度72.63%，结果受类别不均衡性影响较大，部分城市（如西安）分类结果受季节影响较大。

**附：**

**1** 数据丰富流程

