

每日小结

	周一	周二	周三	周四	周五
早	组会 PPT，对比实验补充	PPT，跑代码，多尺度	组会 PPT，对比实验补充，复现	Loss 代码修改，PPT，跑代码	网络图和其他 PPT 的图
中	Loss 想法代码 3，组会 PPT	多尺度，组会 PPT，摘要	摘要，组会 PPT，跑代码	网络图，跑代码	Loss 代码和公式，跑代码
晚	实验补充，PPT，跑代码	代码调参，PPT	整理结果	网络图绘制，跑代码	组会 PPT，跑代码

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: Classifying land-use patterns by integrating time-series electricity data and high-spatial resolution remote sensing imagery

作者: Yao Yao, Xiaoqin Yan, Peng Luo, Yuyun Liang, Shuliang Ren, Ying Hu, Jian Han , Qingfeng Guan

出处: JAG 2022

方法:

基于深度学习构建特征融合神经网络，通过将时序的电力数据与遥感图像耦合，来识别城市土地利用类型。本研究的流程：1) 数据预处理。对遥感影像和时序电力数据进行数据清洗和增强，并将每个样本都进行空间匹配；2) 构建特征融合神经网络（TR-CNN），分析融合时序电力数据和遥感图像的特征对城市土地利用识别的有效性；3) 基于 TR-CNN 识别每个网格地块的土地利用类型。在江西萍乡数据集进行了实验（未公开，电力数据集）

The diagram illustrates the TR-CNN architecture for land-use classification. It starts with 'Multi-Source Data' including 'High-spatial resolution images' and 'Time-series electricity data'. These undergo 'Data Enhancement' and 'Data Cleaning'. The high-resolution images are processed by 'High spatial resolution Image-based ResNet18' with layers: 48x48x64 (kernel: 7x7), 24x24x256 (kernel: 1x1 & 3x3), 6x6x1024 (kernel: 1x1 & 3x3), 3x3x2048 (kernel: 1x1 & 3x3), and 1x1x2048. The time-series electricity data is processed by 'Time-series electricity data-based LSTM-FCN' with layers: Dimension Shuffle, 128x120 (kernel: 8x8), 256x116 (kernel: 5x5), 128x114 (kernel: 3x3), and 1x128. The outputs are combined into a 'Concatenated Feature Vector' (256 + 128 = 256). This vector passes through a 'Softmax' layer (5 classes) to produce the 'Classification Result'. The result is used for 'Land Use Mapping' and 'Land Use Pattern Analysis', showing examples for Residential districts, Commercial zones, Educational facilities, Green land, and Public services. A 'Model performance' section includes 'Model Comparison' (Electricity data: LSTM-FCN, HSR: ResNet18) and 'Ablative analysis'.

启发:

1. 想和这篇做对比实验，没有开源代码，代码在写了。

题目: Multiscale Vision Transformers

出处: arXiv 2021

The diagram illustrates the multi-scale feature fusion process. An input image is divided into four regions (1, 2, 3, 4) and N. These regions are processed by a network to produce multi-scale features. The features are then fused into a single output feature map. The diagram shows the input image, the multi-scale features, and the fusion process.

MViT 引入了多尺度特征金字塔结构，解决了视频识别任务中目标密集型任务。而且参数量与推理速度要比 **ViT** 少很多，在图片分类任务上的表现也比 **ViT** 的要好。**MViT** 是多尺度特征层次结构和 **Transformer** 的结合。**MViT** 有几个通道分辨率尺度块（**channel-resolution scale stages**）。从输入分辨率和小通道维度开始，这些 **stages** 扩展通道容量，同时降低空间分辨率。这创建了一个多尺度特征金字塔，早些的层在高空间分辨率下运行以模拟简单的低级视觉信息，而更深层在空间粗糙但复杂的高维特征上运行。

用pool来减少输入attention的序列长度，做法类似poolformer，都能减少计算量

Figure 3. **Pooling Attention** is a flexible attention mechanism that (i) allows obtaining the reduced space-time resolution ($\hat{T}\hat{H}\hat{W}$) of the input (THW) by pooling the query, $Q = \mathcal{P}(\hat{Q}; \Theta_Q)$, and/or (ii) computes attention on a reduced length ($\tilde{T}\tilde{H}\tilde{W}$) by pooling the key, $K = \mathcal{P}(\hat{K}; \Theta_K)$, and value, $V = \mathcal{P}(\hat{V}; \Theta_V)$, sequences.

启发:

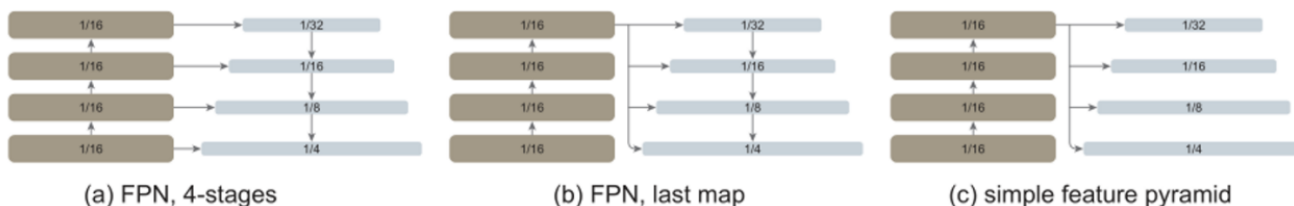
1. 看了代码, 修改了自己的多尺度部分, 也是通过池化做, 增加了反卷积

工作进展

- 1: 阅读文献;
- 2: 组会ppt制作; 网络图绘制
- 3: 补充了大量实验, 放在组会 ppt 里

Visit Model	Image Model	Accuracy	F1 score
ConvTransformer	Resnet18	0.658	0.562
	Resnet50	0.663	0.569
	Resnext50	0.661	0.564
	Alexnet	0.64	0.545
	VGG16	0.656	0.558
	VIT transformer	0.676	0.578
	VIT transformer (Muti-scale)	0.680	0.586

- 4:多尺度模块做出来, 有提升, 在组会 ppt 里汇报

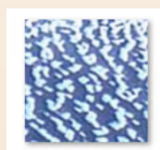


- 5:摘要撰写

- 6: 数据集问题:

数据集

- 1.发现数据集有很多全黑全白图是否需要去除 (小数据集10万张含有2000张左右, 大数据集20000多)
- isprs论文里没写去除全黑全白, 他有没有可能去除了? 精度才超过70%



怀疑数据集靠 高精度土地利用图 分割遥感图像得到
可以去除低信息熵 (但是会去除掉水体部分)

下周计划

1. 网络修改：继续看论文想 idea:
2. self-attention 换多头注意力代码
3. 撰写introduction、related work
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