周报（2023. 2.15-2023.2.22）姓名：孙瑞阳

每日小结

	周一	周二	周三	周四	周五
早	跑模型	跑模型	跑模型 t	阅读文献	信令预处理
中	跑模型	论文阅读	论文阅读	跑代码	信令预处理
晚		论文代码阅读		信令预处理	信令预处理

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

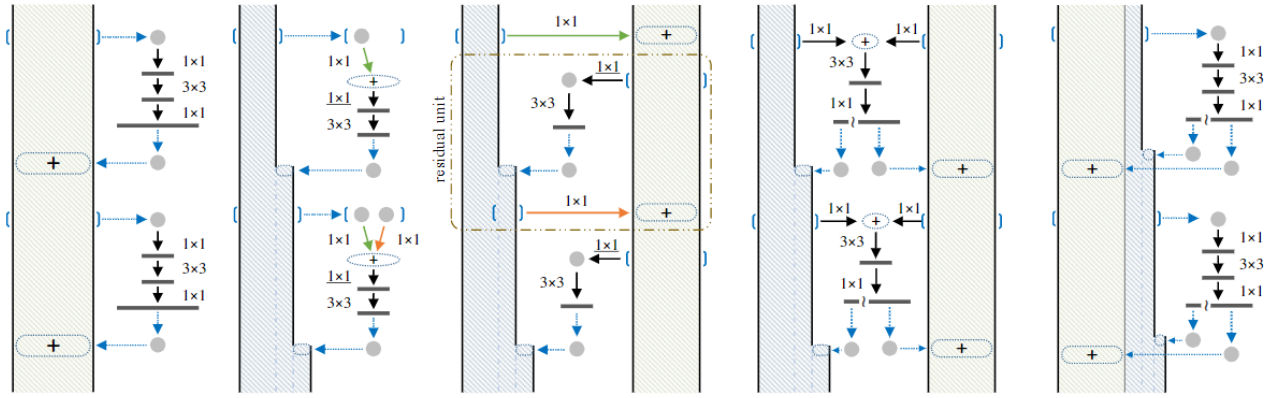
题目： Dual Path Networks

作者： Yunpeng Chen, Jianan Li， Huaxin Xiao， Xiaojie Jin, Shuicheng Yan , Jiashi Feng

出处： Arxiv 2017

方法：

提出了一种用于图像分类的简单、高效和模块化的双路径网络（Dual Path Network /DPN），该神经网络内部连接路径采用了一种新的拓扑结构。通过在 High Order RNN 结构框架下揭示性能最优秀的残差网络（ResNet）和密集卷积神经网络（DenseNet）之间的等价性，发现 ResNet 能重复利用特征，而 DenseNet 能探索新的特征，这两种都对学习一个优秀的表征十分重要。为了获得两种路径拓扑的长处，本文提出了双路径网络（DPN），该神经网络能共享公共特征，并且通过双路径架构保留灵活性以探索新的特征。



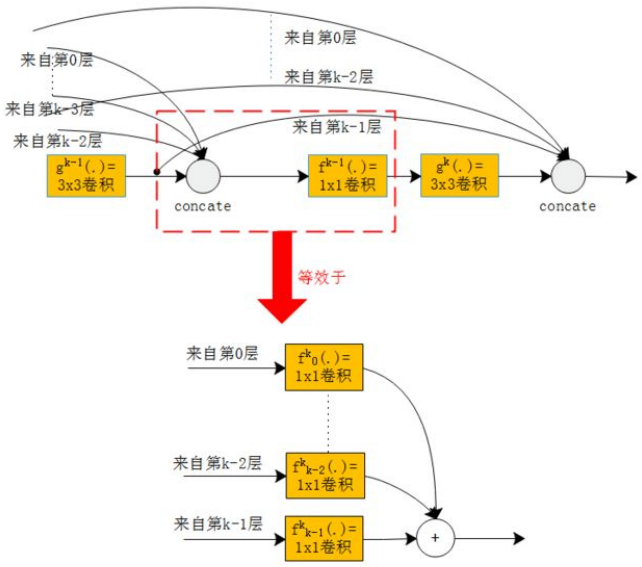
（（a）残差网络；（b）密集连接网络，每一层都可以获取所有先前微模块的输出。这里，添加 1×1 卷积层是为了与（a）中的微模块设计保持一致性；（c）通过共享（b）中层间的相同输出的首个 1×1 连接，密集连接网简并成一个残差网络，（c）中用虚线圈起的长方形标出了残差单元的位置；（d）本篇所提出的双路径结构——dual path architecture——DPN。（e）实现过程中（d）的等价形式，「~」表示一个分支操作（split operation），「+」表示元素级（element-wise）的相加。）

提出的 DPN 网络通过堆叠多个模块化的 block 来实现。本文中，每个 block 的结构设计成瓶颈的风格，1 * 1 卷积然后 3 * 3 卷积，最后再 1 * 1 卷积。最后一个 1 * 1 卷积层的输出分成两部分：第一部分是逐元素相加的残差 path；第二部分是级联的密集连接 path。为了增强每个 block 的学习能力，我们在第二层（3 * 3）中使用 ResNeXt 中的分组卷积。

启发：

1. DenseNet 一个 Block 可以看成图 1 的结构，其中的红框部分是对之前各层的输出在特征维度做拼接，然后做 1X1 卷积。拼接后做 1X1 卷积可以等效为先在各层的直连线上分别做 1X1 卷积（每一条直连线上的 1X1 卷积系数都不同），然后再算术相加。即红框里的结构可以等效为图 1 下半部分的结构，一个 DenseNet 的 Block 就等效于图中的结构（其中 fk-1k、fk-

2k、fk-3k.....都不相同）。所以说 DenseNet 是在不满足 ftk(.)=ft(.)，gk(.)=g(.)时的特殊 HORNN。



2. ResNet 复用了前面层已提取过的特征，除去这些直连的复用特征外，真正由卷积提取出来的特征“纯度”就比较高了，基本都是之前没有提取到过的全新特征，所以 ResNet 提取的特征中冗余度比较低。
3. 而 DenseNet 的 fk-1k、fk-2k、fk-3k.....都不相同，前面层提取出的特征不再是被后面层简单的复用，而是创造了全新的特征，这种结构后面层用卷积提取到的特征很有可能是前面层已提取过的，所以 DenseNet 提取的特征冗余度可能高。
4. 一个有高复用率，但冗余度低；一个能创造新特征，但冗余度高，如果把这两种结构结合起来，DPN 的效果会更好。
5. 论文里和 vgg16、resnet、resnext 等网络结果，比较符合我自己的实验结果，DPN> vgg16 >resnext101> resnet

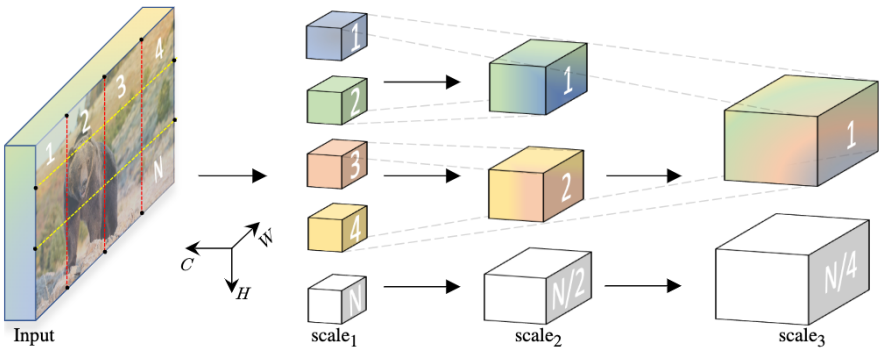
文献2

题目：Multiscale Vision Transformers

作者：Haoqi Fan *， Bo Xiong *, Karttikeya Mangalam *, Yanghao Li *， Zhicheng Yan，Jitendra Malik， Christoph Feichtenhofer

出处：arXiv 2021

方法：



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-resoluation 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。

Multiscale Vision Transformer (MViT)其核心在于增加通道容量，同时减小空间分辨率。

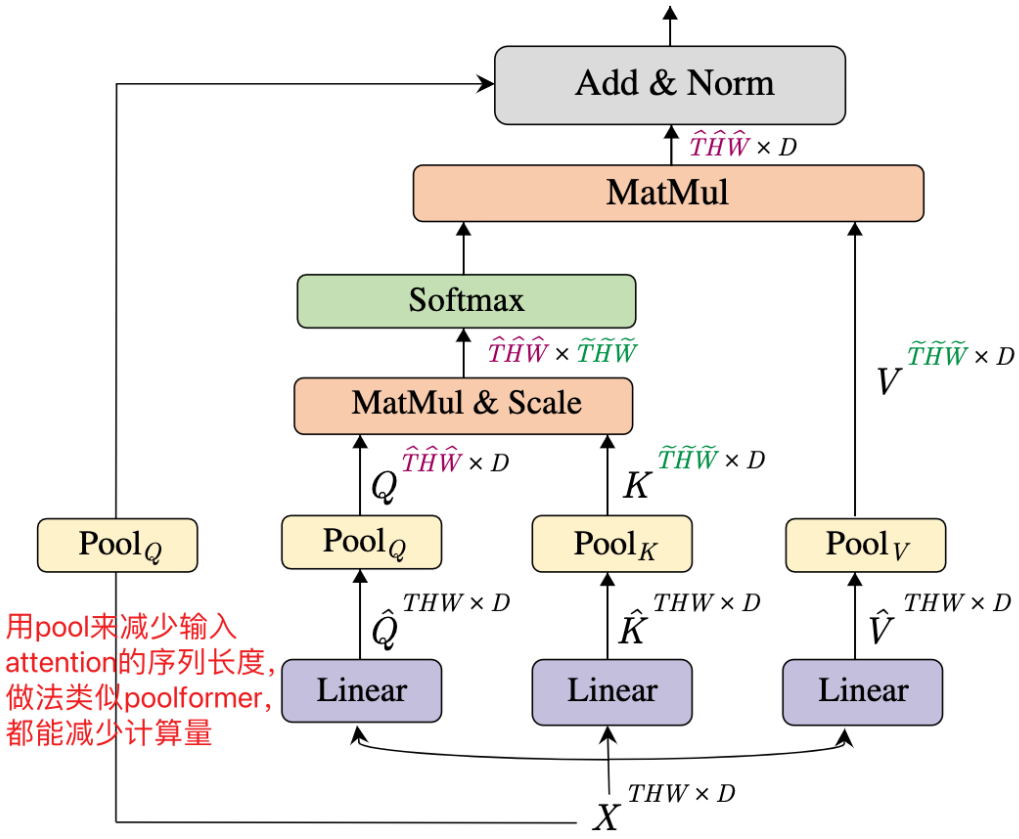


Figure 3. **Pooling Attention** is a flexible attention mechanism that (i) allows obtaining the reduced space-time resolution ($\hat{T}\hat{H}\hat{W}$) of the input (THW) by pooling the query, $Q = \mathcal{P}(\hat{Q}; \mathbf{\Theta}_Q)$, and/or (ii) computes attention on a reduced length ($\tilde{T}\tilde{H}\tilde{W}$)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

启发：

1.利用 pooling 操作实现下采样，**这篇论文的方法简单有效，采用pooling的方式来控制transformer的计算量，值得借鉴作为backbone使用，但是对于效果好的原理还是没能想明白，大概是pooling操作提取了有效信息，丢弃了冗余信息。**

2.不同的 stage 使用不同核大小的 pooling ，这些 stage 连起来就像多尺度金字塔，借鉴之前 CNN 里面的方式，构造特征金字塔，得到多尺度的信息，进行融合学习。。

工作进展

1：阅读文献；

2: 郑州信令数据预处理完成。

3：**实验结果，**补充了九个实验，完成了 cnn+lstm/cnn+bilstm 代码：

模型	准确度	F1指数
resnet50+dpn92	0.675	0.601
transfomer+dpn92	0.666	0.587
transfomer+dpn26(vit b 32) **transfomer+dpn26(vit b 16)**	**0.68** **0.689**	**0.590** **0.598**
resnet50+dpn26	0.660	0.59
resnet101+dpn26	0.657	0.594
resnet101+dpn92	0.653	0.58
resnet101+bilstm毕设	0.668	0.571
transformer+bilstm	0.654	0.554
VGG16+dpn26	0.671	0.598

- 社交网络从bilstm改为cnn+lstm/cnn+bilstm 都跑了

tansformer+CNN LSTM	0.64	0.541
tansformer+CNN BiLSTM	0.626	0.526

- 原来Bilstm输入尺寸（batchsize，182，24）换成周平均（batchsize，7，24）：

tansformer+周平均BiLSTM	0.62	0.52
transformer+**非周平均**bilstm	0.654	0.554

下周计划

1：郑州信令数据输入已经训练好的单模态网络看一下结果。
2: 阅读论文
3: 修改模型