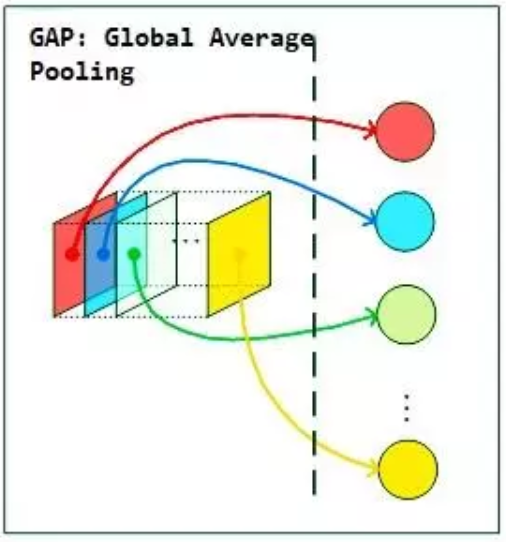
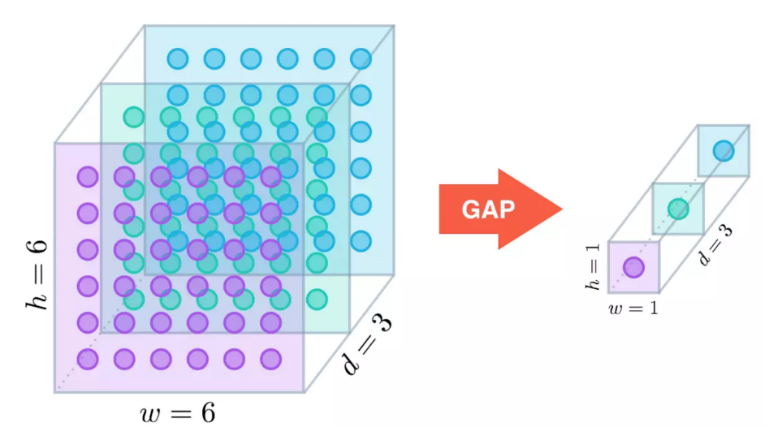
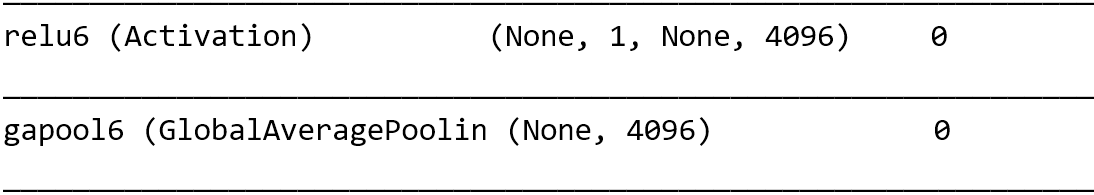
# 说话人识别的话语级聚合

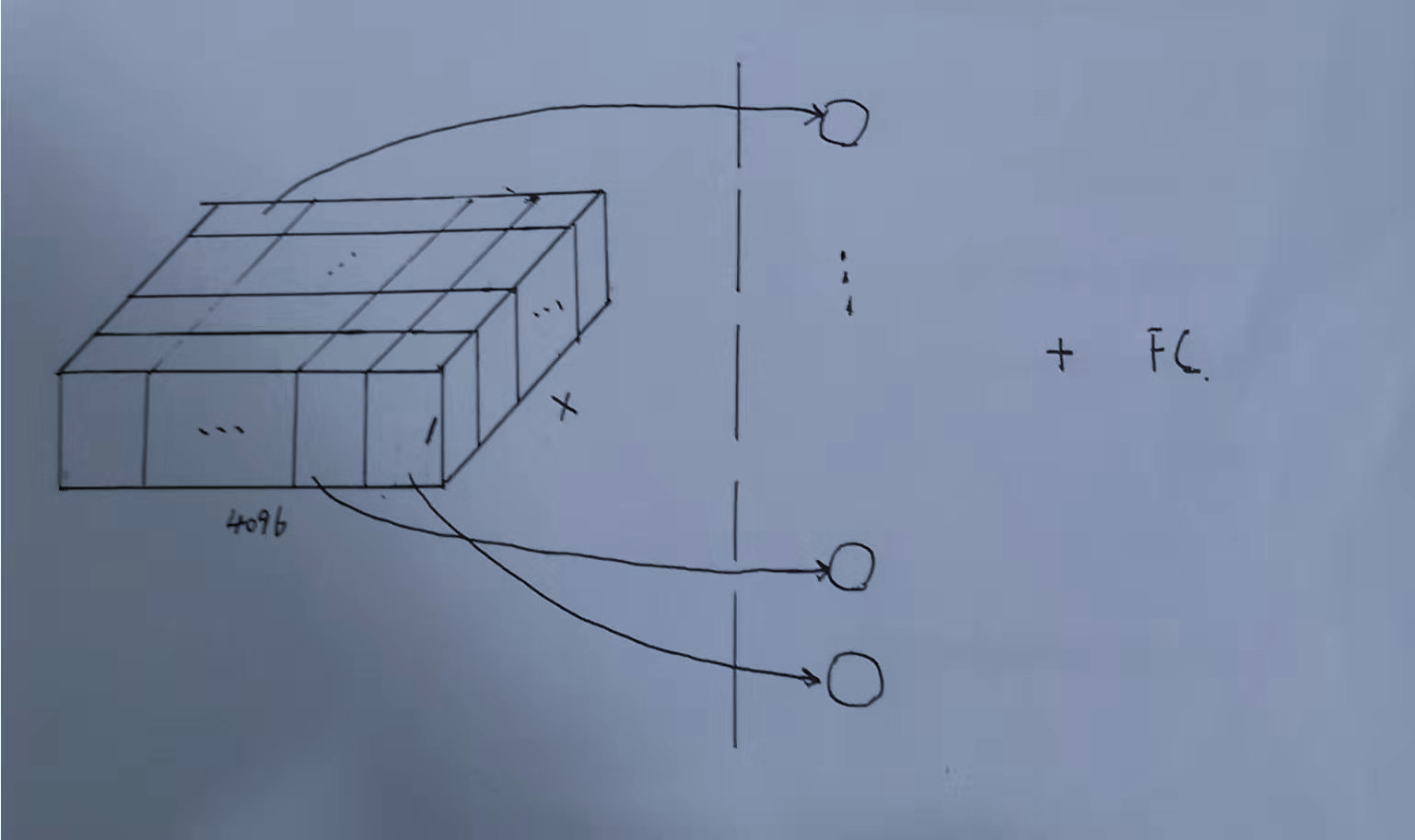
## GlobalAveragePooling

### 可视表示



### Vggvox





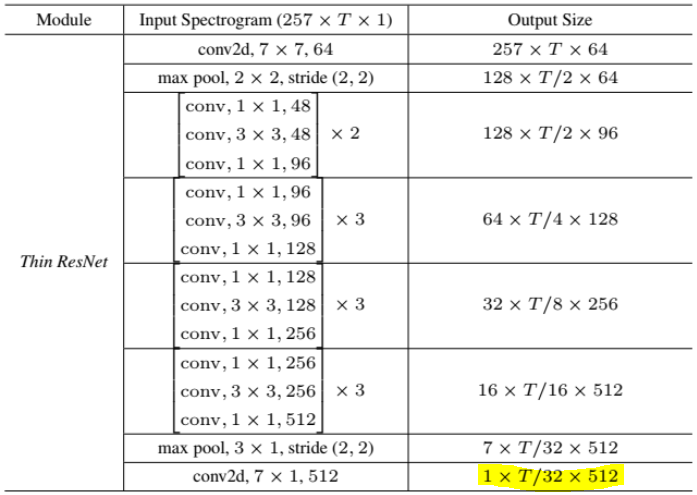
## 话语级聚合

### 主要思想

* 从频谱图中提取特征
* 使用NetVlad将帧级别的特征描述聚合为一个句子级别的向量，替代Gapooling

### 特征提取

* 任意网络 如Vggvox，resNet.
* 论文中使用thin-resNet,参数较标准resNet-34减少 17 million.



### NetVlad聚合

* Vlad
* Vector of locally aggregated descriptors，局部聚合描述向量.
* 一种将图片表示为向量的方法，设计用于大规模图像检索中.
* 算法：

对一幅图像的N个D维特征点 ，聚类得到 K 个聚类中心.

其中，符号函数当属于第k个聚类时，为1，否则为0.

上式计算以为聚类中心的聚类中的特征点的的每一维的值与聚类中心的每一维的值的差的总和.

得到 K\*D 的矩阵 V.

* 特点：

考虑了特征点的每一维的值；

只保存每个特征点到离它最近的聚类中心的距离；

## 资料

Vlad:<http://xueshu.baidu.com/usercenter/paper/show?paperid=ae8cbf70f130b3589efe006e729d8762&site=xueshu_se>