基于深度学习的图像检索算法应用

工作部门：信息技术系

姓名：龚子淳

摘 要：图像是绝大多数生物与外界进行信息交换的重要介质，随着近年来手机等移动设备的普及以及各类网站的兴起，图像数量增长的非常迅速，在此情形下，为了更快速，更准确的进行图片分类，快速识别不同模式、根据早前知识进行归纳、以及适应不同的图像环境并给出图像属于相应分类的概率：在本课题的研究中，应用PatternNet把各个卷积核的组合代表的具体含义筛选出来，把高层混乱的语义信息转化成人能听懂的语义信息，把不同核代表的特征在原图上的位置凸现出来，给予能够理解的分类标准的同时，减轻网络负担。

**关键词：**卷积神经网络 ；深度学习 ；图像分类

# 1.课题的研究现状及研究的目的和意义

## 1.1课题研究的目的

此次课题主要是基于深度学习的图像检索算法的具体实现。首先依据已有的特征分类数据集，对输入计算机中的图像提取特征，并从大量分类特征组成的数据集合中找出和目标物体相似度最高，最为符合的分类特征集合，再将此特征分类数据集合对应分配给所对应的输入图像。

## 1.2课题研究的意义

人工智能方向众多的分支中最为人所熟知的一支就是图像检索，与图像识别有着密不可分的关系。近些年来最令人所熟知的对输入图像进行检索研究的方法就是鼎鼎大名的CNN（中文全称为卷积神经网络）在图像检索，物体判定方面现已成果显著。网络的准确性远超以往传统的特征算法。可若是放在生活实际的环境中，对周边的实物进行图像检索，卷积神经网络的具体效果却并不尽人意。只能大致给出分类，却无法具体展示出其识别的依据。明白在某次预测中，究竟是原图像的哪些部分，分别给出了多少贡献值对于未来更为精确的图像检索是十分必要的。

目前所接触到训练好的网络全部经过了黑盒训练，虽然能够使得卷积核对一类特征感兴趣，但其高层语义信息较为混乱，一个特征可能对应高层的多个卷积核，而一个卷积核也可能对应多个特征。虽然应用经过训练的网络能够检索出相应图片，但依旧只知其然而不知其所以然，无法给出明确的筛选图片的理由。为了减轻网络负担，提高识别准确率，此次试验中应用的网络模式把各个卷积核代表的具体含义筛选出来，把高层混乱的语义信息转化成人能听懂的语义信息，不同核代表的特征在原图上的位置凸现出来，给予能够理解的分类标准的同时，减轻网络负担。

# 2.图像检索的相关技术

## 2.1检索技术分析

图像检索属于人工智能众多分支中最引人瞩目的一支，与图像识别有着密不可分的关系。伴随着科学技术的发展，当今社会已经有了不少相对精确地对输入图像进行检索研究的方法。

(1)立足于文本。在图像数据的检索领域中应用[文本检索](https://baike.baidu.com/item/%E6%96%87%E6%9C%AC%E6%A3%80%E7%B4%A2)的方法，不仅技术发展较为成熟，实现的难度也不高。但缺点是受制于文本信息自身所有的局限性，容易被计算机理解成不同的意思，文本数据迭代更新的速度不高，实在难以面对目前互联网中实时更新的各种图像数据。

(2)立足于图像。目前为止，图像检索领域中已经涌现出多种分析与表达图像特征的方法。相比较来说，针对高层语义信息相比针对中低层次特征的分析来说不仅实现难度要大很多，实现速度也大幅减缓，但能够更好的反应图像在高层语义上的本质特征。

本次实验中主要应用的技术模型PatternNet属于在第二种方法的基础上，针对第二种图像实现速度慢的弱点，以使用反卷积的方式以求用肉眼过滤机器无法识别的噪音干扰，进一步寻找改进的方法。

## 2.2实验技术分析

2.2.1 AlexNet

在互联网时代，信息的流转方式各样，图片是信息的主要载体之一，关于图片分类的算法模型研究层出不穷，AlexNet[2]正是其中最为优秀的代表。在八年前的一次比赛中，AlexNet网络同时对一百多万张包含了上千个种类的图像数据集进行预测，其预测结果的准确程度却远远超出别的算法模型，击败了所有的非卷积神经网络，一战成名。AlexNet的网络结构共有八层，同现在绝大多数的双位数甚至三四位数的卷积神经网络相比，非常简单易懂。AlexNet网络一共有八层。整个网络结构显示如图1所示：

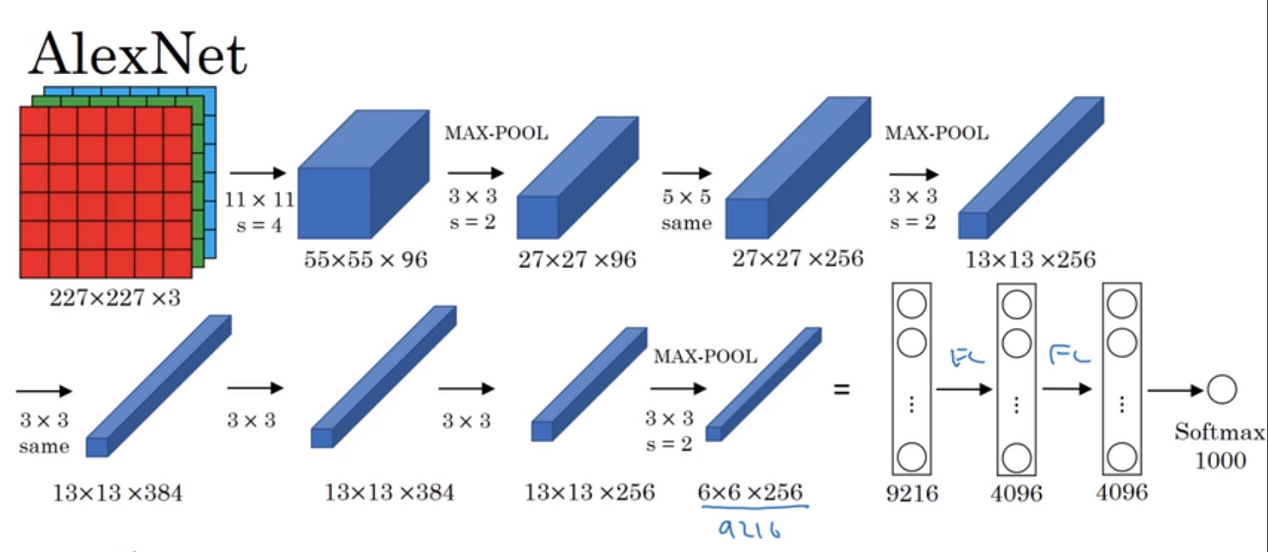


图1完整的AlexNet网络结构图

第一步找到大小为的图片进行输入操作，首先进行卷积操作，使用个大小为的卷积核，步幅设置为，于是对应的图片大小缩小为（由公式（1）可得，大约为原图像大小的） （1）

第二步添加一个宽高f为的卷积核构建最大池化层，卷积核的宽高f为3，步幅s为，卷积层大小变为（公式（1）中卷积公式同样适用于池化操作）,然后让图像执行一个filter=的卷积操作，加入合适的padding（为图片进行边缘补零操作，目的是使卷积操作不改变原图像尺寸n）再然后，其输出的结果即为（使用了256个卷积核）；接着最大池化，使其尺寸n进一步缩小成，再执行与上一步相同的卷积，相同的padding，得到的结果为；再做一次filter=的相同卷积与相同操作，最后再进行一次最大池化，尺寸缩小到了；而，将其展开就可以得到个神经单元，后接全连接层，最后再用函数（归一化指数函数，是二分类函数在多分类上的推广，目的是将多分类的结果以概率的形式展现出来，具体如公式（2）所示）输出，看看它究竟最可能是个可能的对象中的哪一个。

 （2）

其中，函数的输入是从K个不同的线性函数得到的结果，而样本向量 x 属于第 j 个分类的概率为P

2.2.2 反卷积

反卷积可以被近似理解为卷积的逆向操作。反卷积可以将卷积的过程可视化，是一种特殊的正向卷积，先按照一定的比例通过补 0来扩大输入图像的尺寸，接着旋转卷积核，再进行正向卷积。具体与卷积的对比如公式（3）与公式（4）所示：

 （3）

 （4）



# 3.基于深度学习的图像检索算法应用总体设计

## 3.1所需实验平台

硬件配置信息为：i7处理器，8G内存，931GB硬盘；软件配置为：64位的Windows 10家庭版操作系统；实验软件环境为：Anaconda Navigator中整合的沙箱环境；集成环境为：Spyder ；所用编程语言为：Python 3.6。

## 3.2设计流程

IMG_285

图2设计流程图

在此次实验的流程如图2所示，首先需要通过在imageNet，PASCAL VOC等网站中收集实验所需数据，接着采用pytorch库中预训练好的，能在一定程度里识别特征的内置网络，再通过替换操作，将一个特殊的全连接替换掉原有网络全连接层，使得得到能够从中提取出同时具有区别性与频繁性的优秀特征的核的组合频繁性指的是这类特征会在他所属的特征中频繁的出现，就像车很少有尖尖的，车轮这个特征一般只在车上而不是帽子上频繁出现；判别性指的是这类特征只会出现在这个特征里，而不出现在别的地方。比如车在大街上可能有行人，但行人属于噪音。在判别图片里是否有车的时候，图片中行人特征的存不存在对图片中是否有车辆存在没有影响，所以判别车辆时候要找到好的特征，比如车门。再通过反卷积将特征图片中引起卷积核反应的优秀特征所在位置突显出来，并将经过卷积的图片还原成原先大小进行输出，使肉眼即可看出究竟是图片的哪些特殊结构引起了卷积核的反应。这个方法有一个特殊的全链接，能找到能对在目标类别中的某特定模式起反应但是对其他类别不敏感的滤波器集，不需要高质量的数据集，速度也快。

3.2.1问题描述

将通过所选取的网络提取到的特征进行反卷积，反池化操作并将其得到的图像与原来的图像叠加后进行二值化处理或用直接采取叠加对比的方法，使得人的肉眼即可直接看出得到原图像中的哪些卷积核与优秀特征发生了反应，哪些卷积核与噪音发生了反应，从而可以直接将对噪音响应的卷积核剔除，这样不仅加快了识别速率，同时也提高了识别的精准度。

3.2.2试验效果的评价指标

本次实验效果的评价指标通常是用结果图和原图像进行对比，判断特征图中所需要判别的物体和实验结果产生的图像的吻合程度，是否依然存在干扰项。如图3所示，可视化了两个最后一层卷积的卷积核，第二行与第三行分别是一个卷积核，第三行的卷积核只对飞机起作用。通过在不同图像上可视化CNN的卷积层中相同卷积核的响应位置,可以发现卷积核能通过不同的高层信息激活。因为单个卷积核可能响应不止一个高层语义信息，所以很难直接使用一个最拟合的卷积核筛选所需要的物体。换一个角度讲，图片数据中所包含的高层信息也可以同时使多个卷积核响应。因此，可以认为最后一个卷积层中的卷积核只能是用于查找中级潜在信息而没办法代表人能看出来的高级潜在信息。所以，可以把最后一层的卷积核当成中层信息的选择器，而不是让人能够直接肉眼看出来的高层信息的选择器。

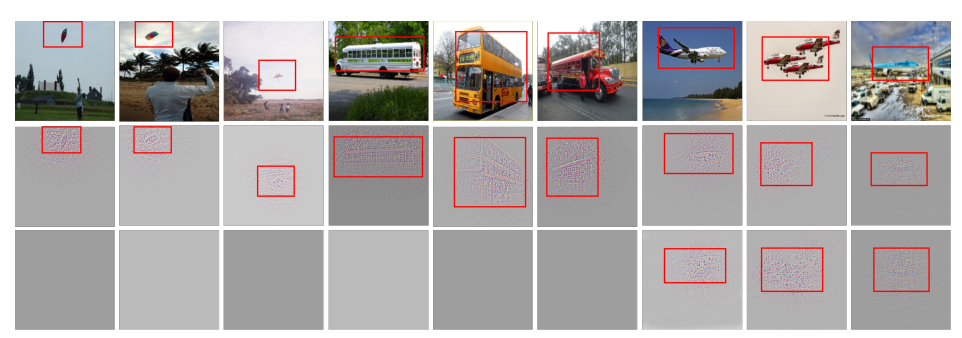


图3图像检索目标效果

# 4.基于深度学习的图像检索算法应用具体实现

## 4.1具体实现

首先从网上下载所需图像数据集，并从pytorch库中调用预训练好的AlexNet网络到该网络的层如图4所示，

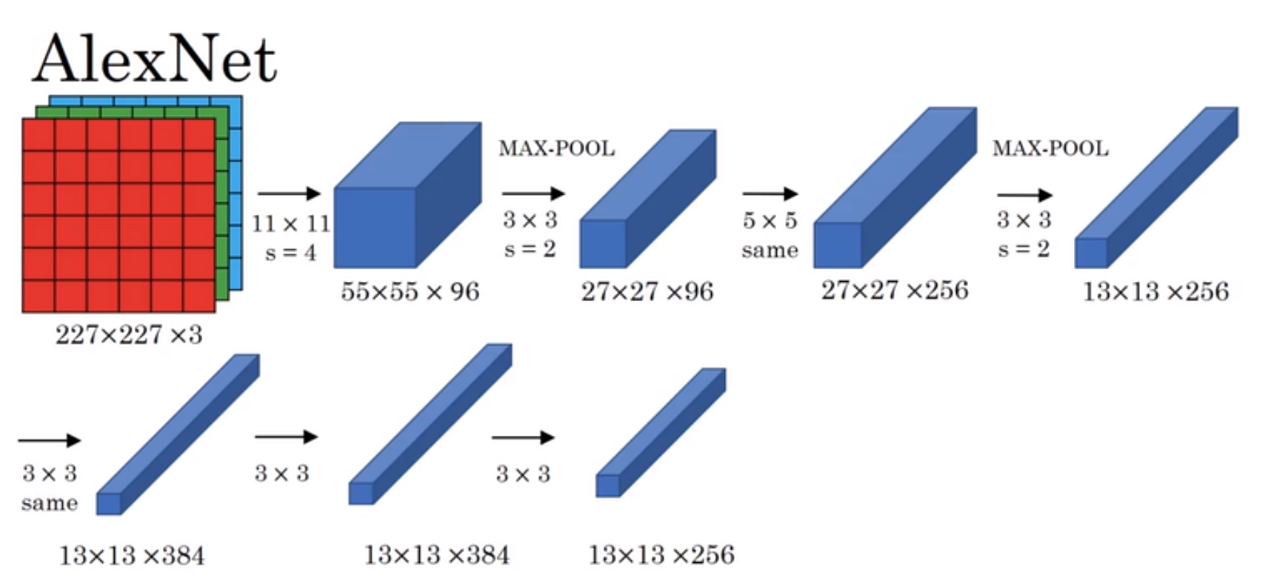


图4实验中完全直接引用的AlexNet部分网络层次

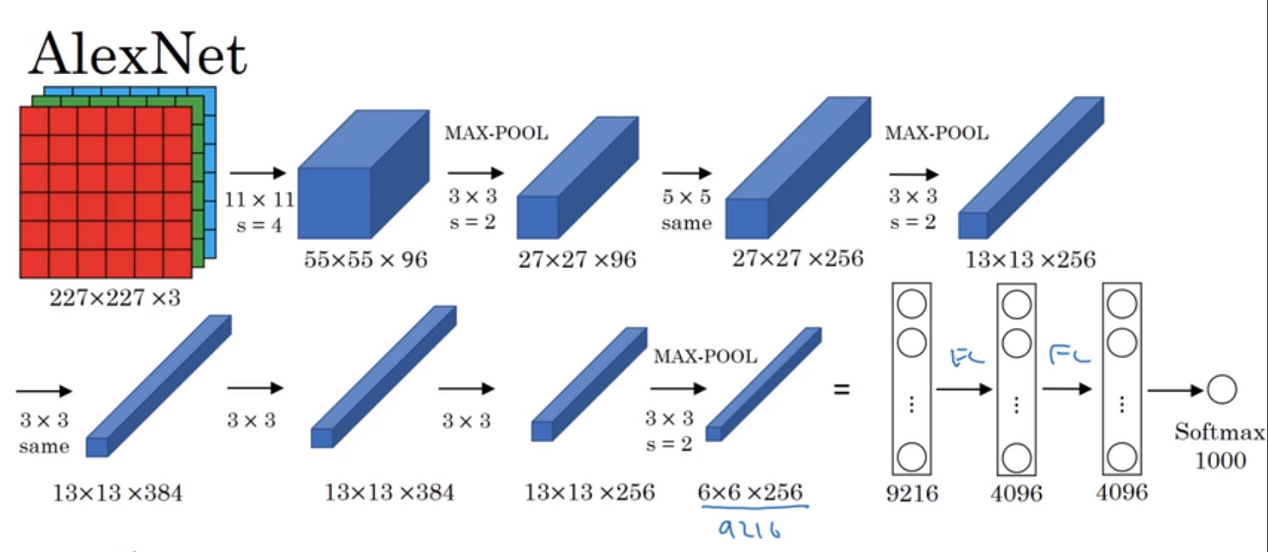
然后进入不属于AlexNet的全连接层，通过最大池化的方式把变成，目的是提取出256个过滤器响应值最大的区域。全连接的第一层有256个神经元，从256个激活值中找180个最大值，进入全连接网络时设为1，其余为0（本应找出一个阈值作为判断0或1的方式，但因为阈值不好找所以以此替代）。之后进行网络训练，加一个全连接层，维度是，把矩阵看成5个卷积核的组合（5个叠在一起，假设第一行代表公共汽车的车窗，第一行里就选几个值最大的权重对应的核当作公共汽车车窗的代表，本行其他的就尽数舍弃）。最后通过这个矩阵的权重，选出核的组合。

图5完整的AlexNet

接着开始反卷积操作（从图5的倒推回原图片），先定义的反卷积网络结构，得到的全连接层的权重，初始化和全连接层FC（）维度相同的全零矩阵，得到一个矩阵，经过处理的FC的权重（从训练好的5个核的集合中，每一个核的集合选出20个相应最大的），得出一个list，里面是所有核的索引（本来有100个就筛选出不重复的卷积核，该运行结果显示有15个重复的，所以提取出剩余的85个不重复的卷积核）。

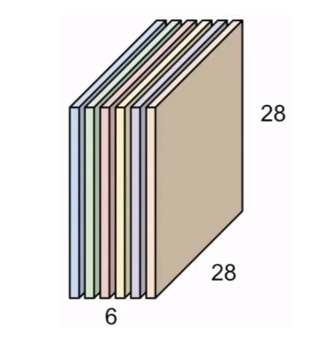


图6示例

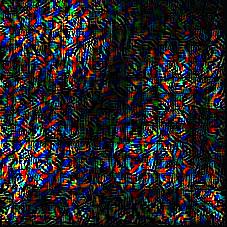
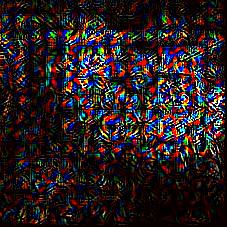
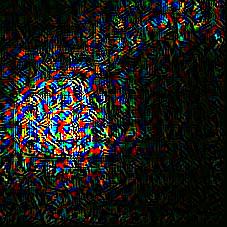
最后将赋值为图6的（图像数据在卷积神经网络的每一个卷积层内以三维形式存在。能够将其看作为无数二维图片的堆叠，其中每一个二维图片都被叫做。图6是一个卷积神经网络中的卷积层，其中，有6个，其中每一个的尺寸都为。又如，我们的彩色照片在计算机当中的存储，就是由三个完成的（RGB模式）；而灰度照片则只有2个），把处理之后得到的新矩阵，赋予（保留上文选中的85个核的响应值，其余归零），接着就是覆盖操作，把反卷积的图当成权重加到原先的图上，这样可以让人眼看出来到底是原图像中的哪些部分引起了相应器的相应。

## 4.2测试结果

通过实验输出的图像，显示不同卷积核在判别的物体和实验结果产生的图像的吻合程度的性能，并可轻松用肉眼看出卷积核到底对输入图像中的哪些特征感兴趣，由此来筛选优秀的特征。



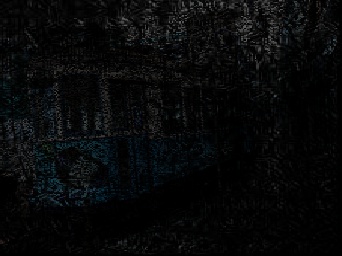
a. 选用的原始图像



b.训练后的网络对原图像卷积后再经过反卷积所得到的图像



c.把反卷积输入图像像素大小当成权重后的图像



d.最终结果，把图像乘上了权重



e.把最终结果进行二值化处理后显示的图像

图7程序运行结果

从图7中的结果b与结果c是程序的中间产物，虽然无法直接看出效果，但结果d与结果e的图像输出是在此基础上得出的，用结果e对比结果a我们可以直观的用肉眼搞清楚经过网络训练的卷积核的组合对图像中的哪些高级语义特征哪里感兴趣，具体是哪些部分引起了卷积核的相应，能一次性的通过网络发现同时具有频繁性和判别性的特征 。图中黑暗部分表示卷积核的响应程度较低，发光部分表示卷积核的响应程度较高，也即是优秀的特征（公共汽车）。

## 4.3实验总结

本次实验研究了我们所用于特征提取的卷积核具体对图片中的哪些特征感兴趣。从此次实验可以看出，我们目前进行特征提取的卷积核有很多不仅仅对所要提取的特征感兴趣，也有可能被图片中的很多噪音所干扰，这点从运行结果的发光部分不仅仅出现在所要识别的公交车上可以看出。但同时也可以看到，通过PatternNet发现的视觉模式既具有频繁性，又具有判别性，把各个卷积核代表的具体含义筛选出来，把高层混乱的语义信息转化成人能听懂的语义信息，能通过哪点判断出来代表的是什么，给予能够理解的分类标准的同时，减轻网络负担。让你明白最高层的哪几个卷积核组合在一起可以分辨具体的公交车所在的图片。

# 5.总结

近年来，越来越多的APP应用了物联网与大数据，出现并融入到人们的生活中，将周边照片上传到网上的同时，肯定也会有越来越多的图像检索需求浮现，换句话说，网络上包含了很多网民打算分享的生活，其中包含着大量重要信息，找出图像检索的具体相应视觉模式在各个方面都发挥着无比重要的作用，我们利用有限的图像数据集，然后得出图像数据所出现的具体特征标志，不仅可以提取同时具有频繁性和判别性的优秀特征的卷积核，提升图片检索速度，还能在某种程度上排除图像中的干扰项，降低图片检索的失误率。

本次实验因硬件与时间限制，采用了卷积神经网络中最为基础的AlexNet，网络深度有限且训练次数比较有限，最后结果虽然已经可以明显看出引起卷积核响应的是哪些图像区域，但却并不十分准确，无法绝对排除噪音干扰，可以看到有时卷积核也对噪音有所响应，但本人坚信，随着知识的积累与程序阅历的提高，将来一定会对此次实验充分理解并改进，更加精准的排除噪音干扰。

# 参考文献

1. 郑莹雪 基于深度学习的图像检索[D].吉林:吉林大学, 2016:4-7
2. Hongzhi Li, Joseph G. Ellis, Lei Zhang and Shih-Fu Chang PatternNet: Visual Pattern Mining with Deep Neural Network ICMR[J],18, June 11-14, 2018, Yokohama, Japan
3. David Bau, Jun-Yan Zhu, Hendrik Strobelt, Zhou Bolei,Joshua B[D]. Tenenbaum, William T. Freeman, and Antonio Torralba. Gan dissection: Visualizing and understanding generative adversarial networks. In ICLR, 2019.
4. Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun[D]. Unified perceptual parsing for scene understanding.In ECCV, 2018.
5. J. Krause, H. Jin, J. Yang, and L. Fei-Fei. Fine-grained recognition without part annotations[J]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5546–5555, 2015.