**文献综述**

课题名称： 观花物语

专 业： 计算机科学与技术

团队成员： 罗雯波、林君宇、李奕萱、郑聪尉

指导老师： 许蕾

2017年10月

1研究背景及意义

随着经济发展和生活水平的提高，人们对花卉的需求不断扩大。截至2014 年年底，我国花卉生产面积已达127.02 万公顷，是世界上最大的花卉生产基地[18]。在享受花卉带来的好处的同时，人们也会遇到一些问题，例如无法辨认花的种类、新手养殖过程中花生长出现各种问题等。目前人工智能技术正在被运用到各个领域去解决问题，已经有相关的技术和产品支持对花卉的图像识别，帮助人们分辨花卉的种类，并且获得相关知识，但市场上的识花App的精度良莠不齐。而且关于如何养花，广大业余爱好者主要依赖于搜索和推送文章，对于养花过程中出现的具体问题，一般依赖于论坛交流，有的也不能及时得到解答。

本项目的意义在于在现有花卉种类识别技术的基础上，进一步提高准确度，并尝试利用人工智能技术更深入地对花卉生长期和健康状态进行识别，为用户提供养殖建议，将识花、养殖指导、论坛交流融为一体，最后做成一个App，既可以让人们在生活中了解到随手拍的花的相关信息，记录生活的足迹，又可以为养花新手解决养花过程中出现的问题，同时也为热爱花草的人们提供一个交流分享的平台。

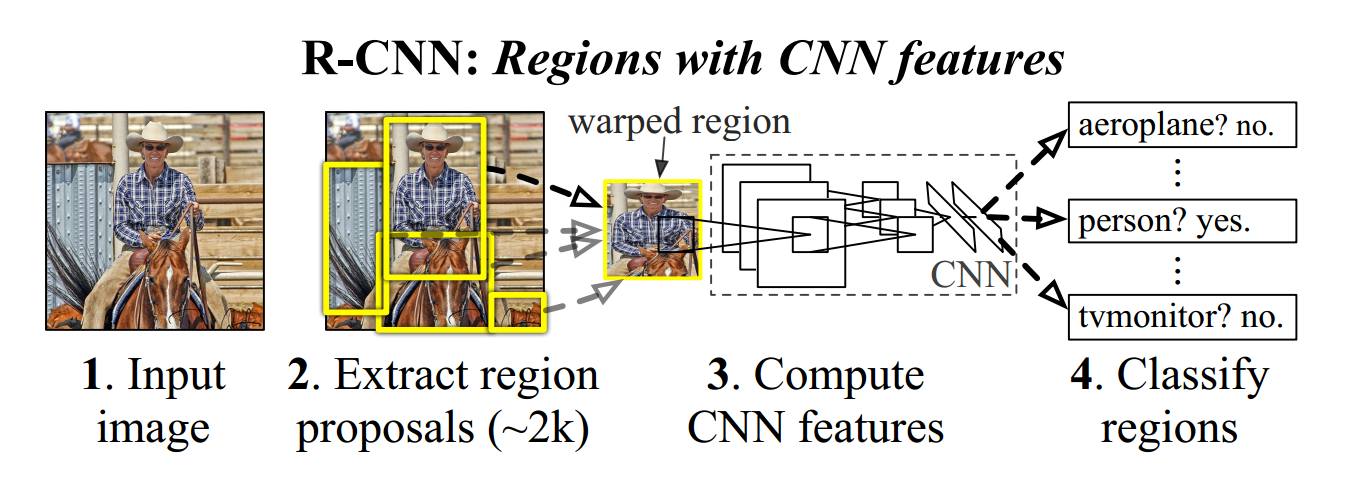
2文献综述

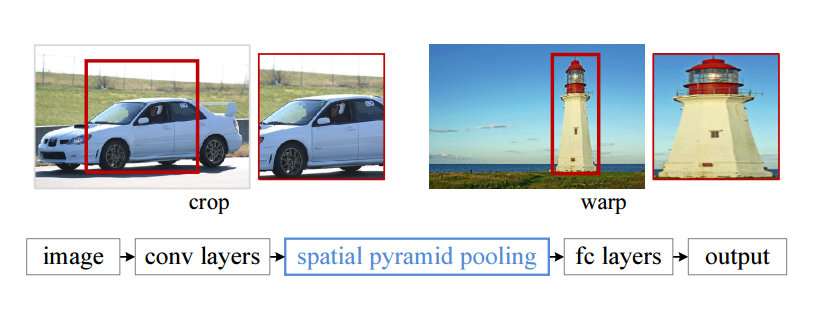
2.1花的定位预处理

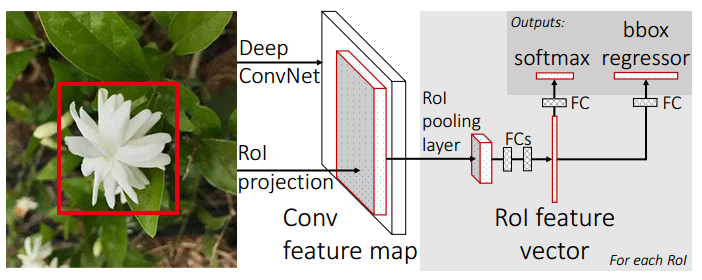
2.1.1基于Region Proposal的深度学习目标检测算法

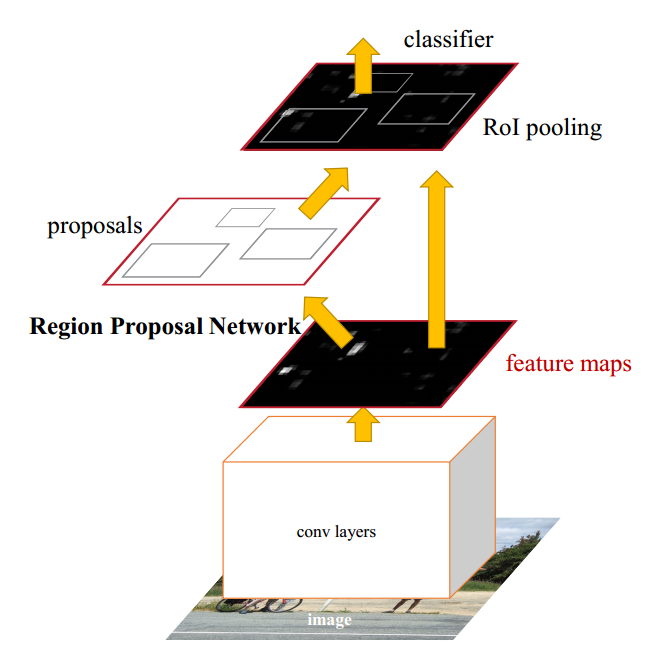
将CNN引入目标检测领域，如R-CNN, SPP-Net, Fast R-CNN和Faster R-CNN等方法，其在PASCAL VOC 2007测试的平均准确率如下：

|  |  |
| --- | --- |
| 目标检测算法 | mAP |
| R-CNN | 58.5% |
| SPP-Net | 60.9% |
| Fast R-CNN | 70.0% |
| Faster R-CNN | 73.2% |

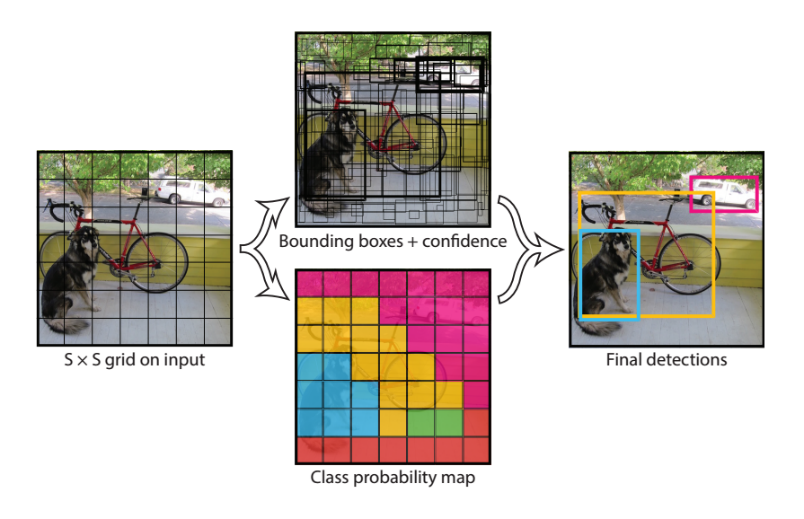
R-CNN方法首先对每一个输入的图片产生近2000个不分种类的候选区域(region proposals)，然后使用CNNs从每个候选框中提取一个固定长度的特征向量(4096维度)，对每个取出的特征向量使用特定种类的线性SVM进行分类，即用识别库进行预训练得到CNN(有监督预训练）后再用检测库调优参数[12]。

相较而言，SPP-Net方法简化了比较耗时的卷积操作，只对整幅图像进行一次卷积运算，使用SPP(Spatial pyramid pooling)将窗口特征图池化为一个固定长度的特征结构，在图像变形情况下仍表现稳定，检测的速度比R-CNN快24-102倍[13]。

Fast R-CNN方法中主要提出了RoI层的结构，以及应用了多任务损失函数(multi-task loss)，包括softmax分类层与Bounding-box Regression回归层，将边框回归直接加入到CNN网络中训练，可进行效果更好的单尺度训练[14]。

Faster R-CNN是一个完全端到端的CNN对象检测模型，使用区域生成网络RPN(Region Proposal Networks)生成建议窗口(proposals)，把建议窗口映射到CNN的最后一层卷积feature map上，通过RoI pooling layer使每个RoI生成固定尺寸的feature map，利用探测分类概率和探测边框回归对分类概率和边框回归联合训练[15]。

2.1.2基于回归方法的深度学习目标检测算法



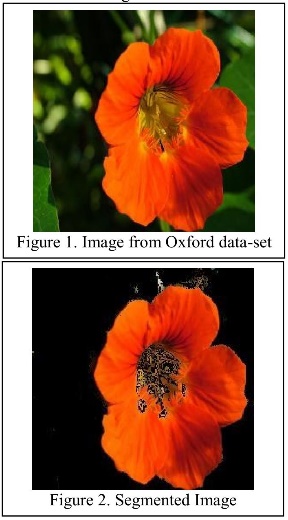
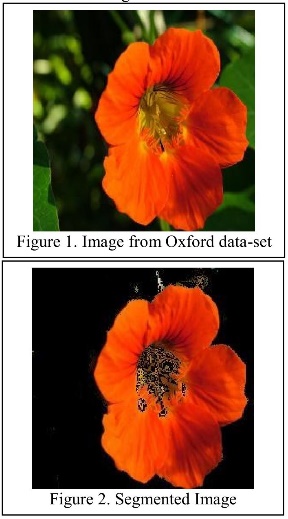
基于回归方法的深度学习目标检测算法有YOLO, SSD以及YOLO2等，定位准确性比Faster R-CNN低但其识别效率较高，可进行实时检测。比如在YOLO方法检测模型中（如上图）,直接在输出层回归bounding box的位置和bounding box所属的类别，把目标检测问题转化成一个回归问题，卷积层进行图像特征提取，而全连接层用于后续类别和目标位置的回归预测[16]。

2.2花的种类识别：

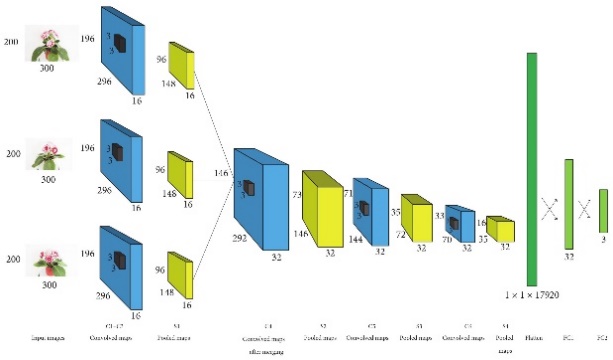
花卉分类本质上是图像识别，过去有DPM，HOG，SIFT等算法用于特征识别，而现在在图像识别领域卷积神经网络（CNN）被更广泛使用。CNN具有高鲁棒性且无需人工处理图像特征。如同生物神经网络，CNN由一系列神经元构成，每个神经元接收数据进行处理，将结果传递给下一神经元，并通过最小化损失函数来对网络进行优化，以提高预测正确率。训练CNN需要相应的数据集，以花卉分类为例，则需要准备大量花卉图片和对应的花卉类名。

对于一般图像识别，Alex Krizhevsky，Ilya Sutskever，Geoffrey E. Hinton（2012）训练了一个6千万参数，650,000个神经元的CNN（AlexNet），包含5个卷积层，若干最大化池和3个全连接层。为了减少过度拟合，提出了dropout的策略。在拥有1000不同类，120万张高分辨率图片的LSVRC-2010训练集中获得了top-1 62.5%和top-5 83%的正确率[1]。

由于花卉形态颜色多样的特殊性，花卉分类的难度要高于一般图像分类。Ayesha Gurnani，Viraj Mavani（2017）对牛津大学的102类花卉图片集进行了Segment预处理[5]，并对比了CNN模型AlexNet和GooLeNet[4]在该训练集上的效果，top-1正确率前后两者分别达到了43.4%和47.2%[2]。



为了提升花卉识别的准确性，我们提出了多图识别和视频识别的思路：除常见的单图识别外，为用户提供上传同一朵花的多张不同角度图片，上传一段记录了花卉不同角度的短视频功能。

 多图识别方面，Yu Sun等在一项关于使用神经网络进行花卉评级的研究中设计了一种多输入的CNN。该CNN模型获取同一株植物的三张不同方位图片作为输入，三个输入分别进入不同的卷积层然后由池化层聚合，再经过一系列层后获得一个预测。经比较多输入的CNN比单输入的CNN在评级上正确率有明显提升，而且多输入而非多次单输入有效减少了预测次数，提高了效率[3]。

视频识别方面，Andrej Karpathy等使用CNN在YouTube-1M数据集上实现了63.9%的准确率[5]。但目前的视频识别主要针对Large-scale的物品分类，如车，房子，足球等，与之相比，花卉识别是更细致的分类，因此视频识别在这方面并不能有很好的应用。但我们认为视频识别仍可以作为多图识别的一种辅助手段，例如采取Single Frame，Late Fusion，Early Fusion，Slow Fusion等模式从花卉视频中抽取关键帧，作为多图识别的输入，又例如利用视频识别中的一些定位、分割技术对关键帧进行处理。

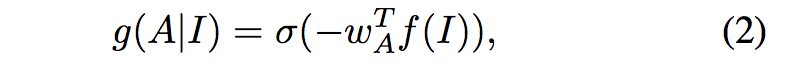
2.3花的状态识别

许多计算机视觉的工作专注于不同物体的识别[6、7]、不同场景的识别[8、9]以及不同材料的识别[10]，但是在一种物体分类中对不同状态的识别也是非常有意义的，对于苹果来说，一个是青涩的苹果，一个是成熟的苹果，如果可以对青涩和成熟进行分类的话，有助于人类对物体属性的深一步理解和利用。对花也是如此，主要借鉴CVPR2015的论文Discovering States and Transformations in Image Collections[17]来进行分析的。

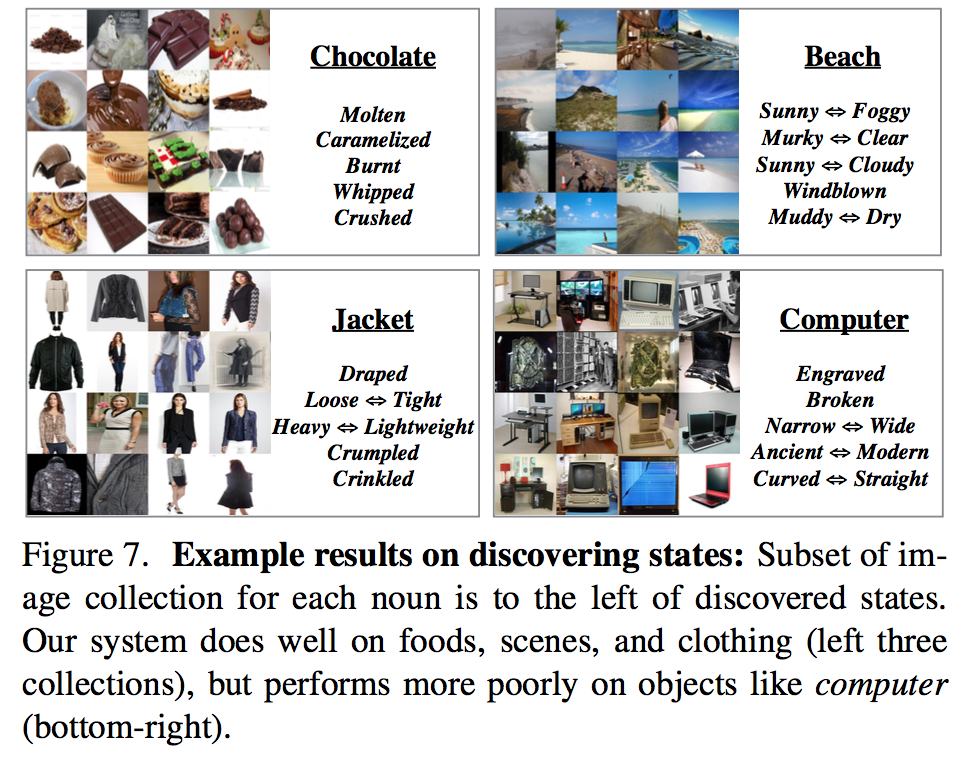
对花品种的识别属于对名词的一个分类，而对于在一个花的类别里判断其生长周期、健康状况以及品质好坏是对于形容词的一个分类，基于不同种类的物体他们在同一状态有类似的特征这一点，比如融化的巧克力、融化的奶油、融化的糖，它们虽然颜色、材质都不相同，但是都有光滑的性质。可以通过对一组物体的一个状态的学习，例如（牛肉，生的）（猪肉，熟的），训练出一个模型用于预测羊肉的状态，尽管在训练时根本没学习过羊肉的图片也可以用于预测。

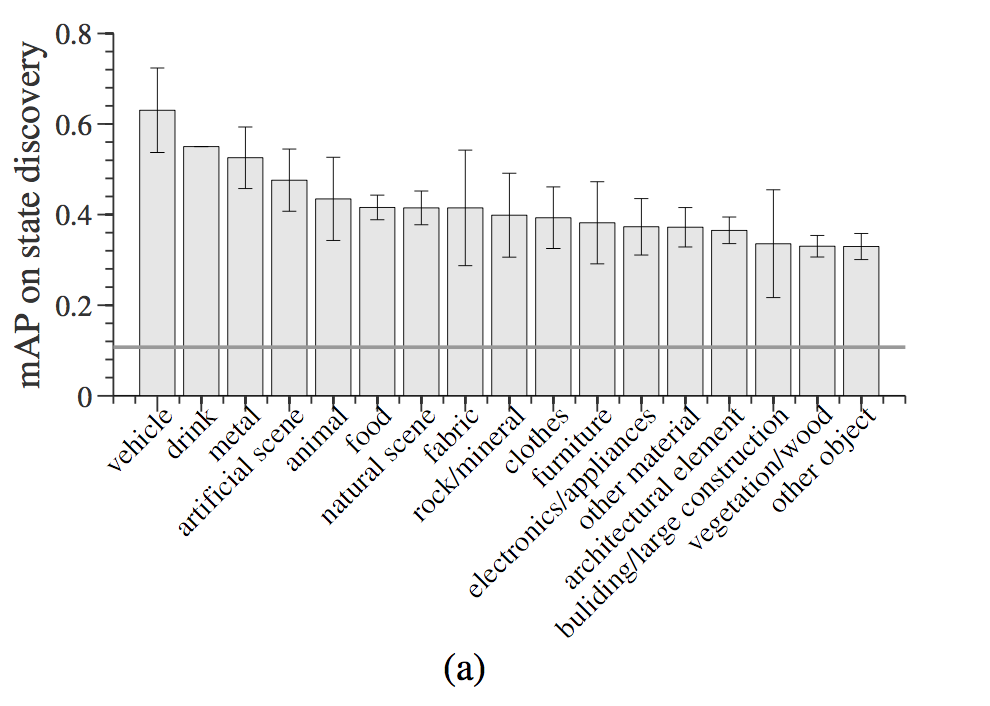
为了完成这个目标，我们有两个主要任务，第一个是建立(noun, adjective)的数据库用于训练模型，第二个是状态分类。

建立数据库分两个步骤，第一是形容词和名词的选择，选取m个名词构成集合N， n个形容词构成集合A，对其进行搭配，利用微软提供的Web N-gram Services评估m\*n个结果中合理搭配的二元组，最后生成M个(noun, adjective)的组合作为图片的标签；论文[11]选取了115个形容词和249个名词，训练结束后，大概每个名词有9个相关的形容词与之匹配成二元组。第二是图片的搜集和选择，在搜索引擎上通过搜索(noun,adjective)二元组获得图片并人工去除噪声图片。以下是论文[11]中最终得到的数据集中三个名词的举例。

要完成状态分类这个任务，首先要把数据库中所有的标签中的名次去掉，只剩下形容词，对于A中的一个形容词状态a，则将全部图片根据状态a进行分类，与状态a相关的列为positive，反之列为negative，这样一幅图片I相对于状态a的得分就可以被式子（2）所表示，其中f(I)是I的一个CNN特征向量(Caffe Reference ImageNet Model layer fc7 features) [11]，w是由logistic regres- sion模型训练出来的，σ 是sigmoid函数。

一张图片可以有不同的状态，可以根据其不同状态的分高低来判断哪个状态是其主要的状态。

训练结果在食物、场景和衣服上做得很好，但是在像电脑这样的物品上做得就不太好，Figure7，(a)展示了对不同种类的名次进行状态分类的结果评估。

3发展现状

3.1用户需求

目前，会使用花草App的用户分为三类：对花草不太了解但很感兴趣的人；喜欢花并且喜欢养花的人，而这类人又可以分为新手和经验丰富的人。对于第一种客户而言，会对平时生活中出现的不认识的花很好奇，可以识别花种类的App可以帮助用户快速获得该花的相关知识，并一键生成美图记录生活中遇见的每一朵花；对于一个养殖新手来说，当花生长出现问题时会手足无措，通过描述花的性状进行搜索，很多时候是无法将具体情况和网上的搜索结果匹配的，因此可以分析花生长状况的App可以快速诊断花的问题，并为用户检索相关解决方法；对于资深花草爱好者来说，乐于欣赏别人的花草也乐于分享自己的心得，因此一个社交平台性质的App会满足这类用户的需求。

3.2现有软件

目前与花卉相关的App已有多款，主要分为两个类别，一种偏重于花卉识别，其中精度比较高的是“形色”，另一种偏重于花卉知识分享、论坛交流，比较流行的是“绿手指”。

3.3主要功能

“形色”在花卉种类识别方面的功能比较强大，准确率较高。在识别出花卉种类后，还会显示相关的百科知识，以及提供了一键生成美图的功能，但是图片的美化效果比较差。“绿手指”的识别功能则不强，主要的功能是论坛分享，给广大花草爱好者提供了一个分享动态、交流经历以及养殖求助的平台。

同时这两款App都有花的种类以及养殖知识的普及文章推送，其中“形色”的实现更好，资源比较丰富且对文章进行了分类，但是缺少搜索功能，不便于快速检索。下表是具体对比。

|  |  |  |
| --- | --- | --- |
| 功能 | 形色 | 绿手指 |
| 花卉种类识别 | 有且准确度较高，有取景框，会给出三种最可能的花的品种，如果没有识别出来就会建议人工鉴定。 | 有识别功能但是很难用，一点进就闪退，一直没成功过。 |
| 论坛 | 1.可以把识花结果发在发在一个类似地图状物，用户可以看到自己附近都有哪些用户上传了识花结果。  2.在“花间”里又一个话题采取帖子机制，发起一个话题大家跟帖。 | 1.求助花的种类或者花出现问题了  2.分享动态，类似于朋友圈，可以对感兴趣的人加关注进行留言交流。 |
| 百度百科信息 | 诗词，典故，文化，价值，养护 | 因为没有识花功能，所以这一点没体验到 |
| 美图生成 | 体验后感觉制作模版比较刻板而且图片感觉有点变形，有优化提升的空间。 | 没有图像处理，直接是图片加文字，属于微信朋友圈格式。 |
| 用户个人空间 | 记录识花时间地点（足迹） | 有一个“花园”，里面有养花的日程表，花成长历程的图片 |
| 花的种类以及养殖知识普及（类似于微信公众号） | 分门别类对文章进行归类，有关与生活的有关与户外的一系列关于花的知识，还有对每个城市赏花景点的推荐，但是缺少搜索功能。 | 与形色类似，但是没有形色做得好。 |

3.4不足之处

现有产品在花卉种类识别的功能上已经有了较好实现，但是对业余养花爱好者没有提供较好的技术支持。在养殖方面遭遇问题时，人们通常只能自己查看资料寻找对策，但是往往毫无头绪，不知道从什么方向去找；或是拍照发到论坛上请求他人帮助，而这种求助通常不能得到及时回应。如果软件产品能够对花卉的生长状态进行一定程度的识别，一定能够给养花爱好者提供有价值的建议，帮助其找到问题原因并解决。

另外，现有产品的花卉美图生成的功能较差，不能满足用户的审美需求。同时对花卉知识的检索主要是以种类为关键词进行检索，没有提供较全面的检索功能。

4参考文献

[1] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. (2012). ImageNet Classification with Deep *Convolutional Neural Networks*

[2] Gurnani, Ayesha & Mavani, Viraj. (2017). *Flower Categorization using Deep Convolutional Neural Networks*.

[3] Yu Sun, Lin Zhu, Guan Wang, and Fang Zhao, *“Multi-Input Convolutional Neural Network for Flower Grading,”* Journal of Electrical and Computer Engineering, vol. 2017, Article ID 9240407, 8 pages, 2017.

[4] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, *“Going deeper with convolutions”*, Proceedings of the IEEE CVPR 2015

[5] Mavani, Viraj, Ayesha Gurnani, and Jhanvi Shah. *"A Novel Approach for Image Segmentation based on Histograms computed from Hue-data."* arXiv preprint arXiv:1707.09643 (2017).

[6] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical im- age database. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009. 1, 2

[7] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Ob- ject Classes Challenge 2007 (VOC2007) Results. 1, 2

[8] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Tor- ralba. Sun database: Large-scale scene recognition from abbey to zoo. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 1, 2

[9] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning Deep Features for Scene Recog- nition using Places Database. *Advances in Neural In- formation Processing Systems*, 2014. 1, 2

[10] L. Sharan, C. Liu, R. Rosenholtz, and E. H. Adel- son. Recognizing materials using perceptually in- spired features. *International journal of computer vi- sion*, 103(3):348–371, 2013. 1, 2

[11] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. *arXiv preprint arXiv:1310.1531*, 2013. 2, 5

[12] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition.

[14] R. Girshick. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, 2015. 2, 6, 8

[15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.

[16] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection.

[17] Phillip Isola∗,Joseph J. Lim∗ and Edward H. Adelson Discovering States and Transformations in Image Collections In CVPR 2015

[18] 白晓琦,苏庆,单会霖. 我国花卉业现状与发展趋势[J]. 现代农业科技,2015,(19):164.