**C9 ZFQ队 设计方案**

1. **算法方案介绍**

将赛题看作“目标检测”问题，核心算法参考YOLOv2，链接<https://pjreddie.com/darknet/yolo/>。

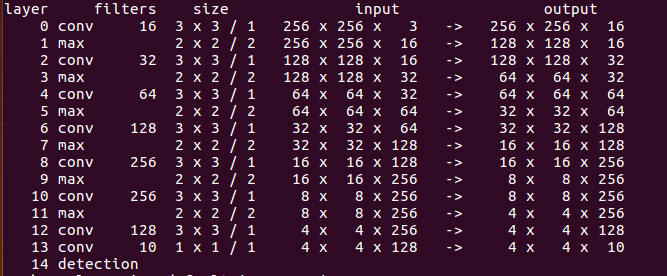
**1.1 算法概要**

1.从视频流中获取一张当前时刻的图像；

2.计时开始，首先将输入图像resize到256\*256\*3, 再输入已训练好的CNN，结构如下，得到4\*4\*10的检测结果， 即在特征图的4\*4个cell中，每个cell预测一个anchor box，每个anchor box预测一个bbox(BoundingBox)的相对位置(x,y)、相对大小(w,h)、置信度(o)以及检测到的手势属于每个类别的概率，共10个值；

3.对4\*4个检测结果采用nms算法进行压缩排序，选择置信度最大的bbox作为最终检测结果，得到bbox的四个坐标值和属于每个类别的概率，计时结束；

4.打印处理这一帧图片所花时间，在原图上标记检测结果，输出到视频流并显示。



检测网络hpnet7结构

**1.2 数据集和训练**

我们采集了2403张手势照片，分别来自9个人（7女2男），左右手随机，每个类别数量相似，统一resize成448\*448\*3的彩色jpg图片格式，使用labelImg标注（https://github.com/tzutalin/labelImg）。训练时，我们采用带动量项的SGD方法，weight decay恒定为0.0001，momentum恒定为0.9，batch size为64 ，初始learning rate为0.1, 分别在总的batch数的30%和60% 时除以10，初始权值通过高斯函数随机初始化。我们首先使用448\*448\*3的输入训练了40000次，再将输入变小为256\*256\*3, 使用得到的权值文件fine-tune了60000次，在训练过程中，也采用了标准的data augmentation，包括随机水平翻转，保证了左右手同一手势的鲁棒性，随机旋转0-15度，饱和度、色调和曝光值随机变化等。

1. **测试数据及验证**

算法并没有特别的测试数据及方法，采集的所有图片均用于训练，实时检测时网络会打印对应预测框的概率值并直接将检测结果标记在输出图片上，我们通过肉眼判断网络输出的预测bbox和groundtruth之间的IOU以及预测类别与真实类别是否相同。

而在训练数据集上的评估方法与结果也只能作为参考，在训练时我们会得到每一次迭代的平均IOU和类别预测正确的概率值，将几种算法的结果列在下表，仅作参考。几种算法的详细结构见附录。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Operations per image（billion） | Avg IOU | recall | FPS |
| yolov2 | 27.5 | 0.86 | 1.0 | - |
| tiny-yolo | 6.95 | 0.85 | 1.0 | - |
| tiny-yolo2 | 3.74 | 0.8323 | 0.98 | **-** |
| hpnet2 | 2.15 | 0.80 | 0.99 | 0.555 |
| hpnet3 | 1.28 | 0.77 | 0.96 | 0.564 |
| hpnet4 | 1.21 | 0.77 | 0.95 | 0.578 |
| hpnet5 | 0.89 | 0.76 | 0.95 | 1.33 |
| hpnet7 | 0.53 | 0.75 | 0.93 | **1.78** |

1. **总结**

1. 基于deep learning思想，参考yolov2算法，我们采用了更小的CNN，参考resnet等先进网络的实现，采用3\*3的卷积核大小，并在每一次卷积后将特征图下采样为一半大小，再成倍增加卷积核数量，共执行6次池化，8次卷积，将原始图像尺寸缩小64倍，最后采用带有10个卷积核的检测层直接得到end-to-end的检测结果;

2. 网络只有卷积和池化层，没有全连接层，每个anchor box只预测一个bbox，网络结构在保证准确率的前提下做了最大的压缩;

3. 将原始图像resize到256\*256, 这一改进大大增加了我们网络的运行速度，比起448\*448的输入快了三倍，但是输入图像再变小的话，会失去很多细节信息导致手势距摄像头大于1m时检测效果不好，故我们选择将输入resize为256\*256；

4. 训练方法上，我们首先使用易训练的高分辨率网络做预训练，再在低分辨率网络上做fine-tune，比起直接在低分辨率网络训练，大大减小了训练的难度，也增加了网络的准确率;

5. 由于共产生4\*4个预测bbox，所以可以同时检测多个目标，理论上最多可达16个;

6.基于deep learning的思想，我们的算法通过让网络自己学习辨别标注数据的特征，再用来在未知数据上测试，并在数据采集过程和训练过程中尽量制造不同的环境来扩充数据集，都能够提高算法的泛化能力，使得算法的鲁棒性很好。我们相信如果进一步扩充数据集进行更好的超参数选择和fine-tune会获得更好的鲁棒性与准确率。

1. **附件**

代码fork自https://github.com/pjreddie/darknet ，自己主要修改src/demo.c用于检测并显示，完整的见gitlab, https://219.139.34.186/zhaofenqiang/Gesture\_C9

backup文件夹为训练好的模型

cfg文件夹为网络结构配置文件

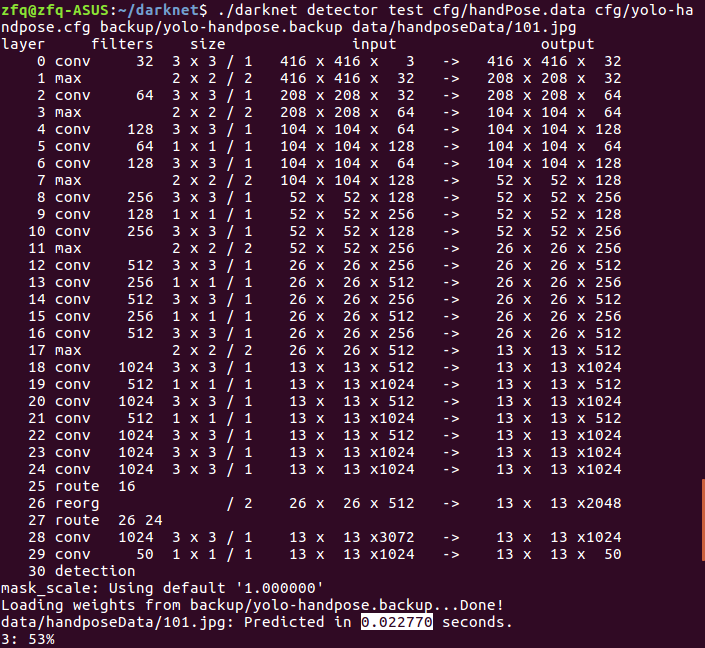
src文件夹为C语言源代码

data文件夹为demo用的一些图片

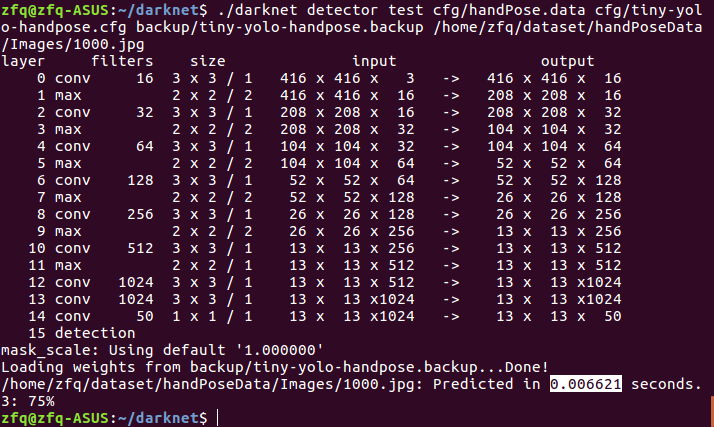
数据集由于国内还是百度网盘方便，所以分享一个百度网盘的链接，这是原始图片数据集，resize后的正在上传，后面要求的话可以补上。

**附录**

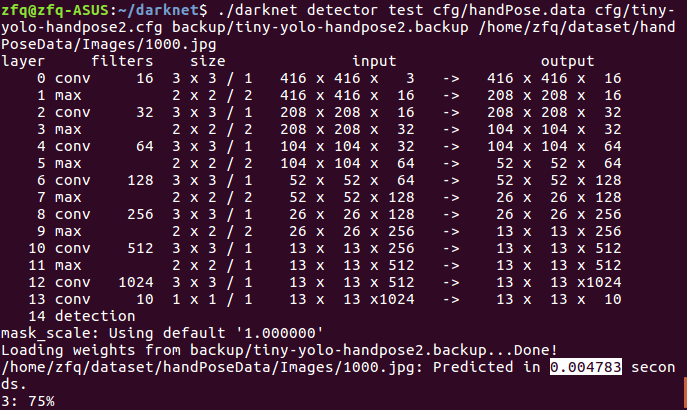
yolov2网络结构



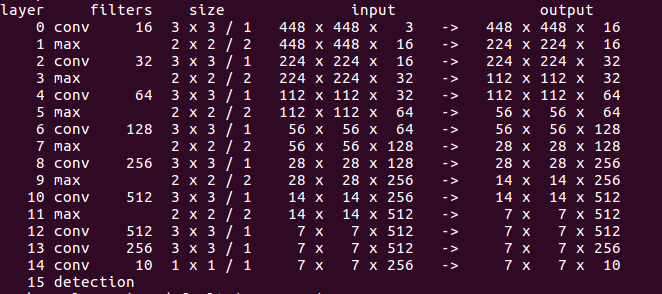
tiny-yolo2网络结构



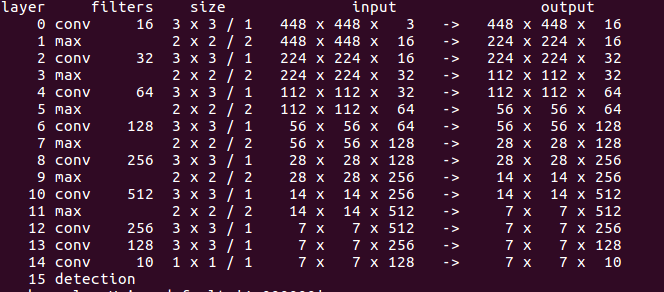
tiny-yolo网络结构



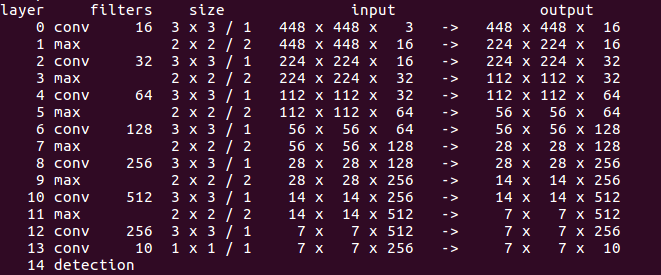
hpnet2网络结构

****

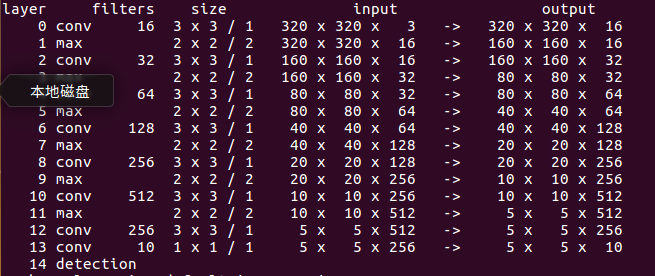
hpnet3网络结构

****

hpnet4网络结构



hpnet5网络结构

****

hpnet7网络结构

